REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the ti and maintaining the data needed, and completing and reviewing the collection of information. Send comments re information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for inf 1204, Arlington, VA 22202-4302, and to the Office of management and Budget, Paperwork Reduction Project (0704-0

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE December, 1995	3. REPORT Final

4. TITLE AND SUBTITLE
USAF Summer Research Program - 1995 Summer Faculty Research Program Final Reports, Volume 5C, Wright Laboratory

5. FUNDING NUMBERS

6. AUTHORS
Gary Moore

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Research and Development Labs, Culver City, CA

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR/NI
4040 Fairfax Dr, Suite 500
Arlington, VA 22203-1613

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
Contract Number: F49620-93-C-0063

12a. DISTRIBUTION AVAILABILITY STATEMENT
Approved for Public Release

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The United States Air Force Summer Faculty Research Program (USAF- SFRP) is designed to introduce university, college, and technical institute faculty members to Air Force research. This is accomplished by the faculty members being selected on a nationally advertised competitive basis during the summer intersession period to perform research at Air Force Research Laboratory Technical Directorates and Air Force Air Logistics Centers. Each participant provided a report of their research, and these reports are consolidated into this annual report.

14. SUBJECT TERMS
AIR FORCE RESEARCH, AIR FORCE, ENGINEERING, LABORATORIES, REPORTS, SUMMER, UNIVERSITIES

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239.18
Designed using WordPerfect 6.1, AFOSR/XPP, Oct 96

UNITED STATES AIR FORCE

SUMMER RESEARCH PROGRAM -- 1995

SUMMER FACULTY RESEARCH PROGRAM FINAL REPORTS

VOLUME 5C

WRIGHT LABORATORY

RESEARCH & DEVELOPMENT LABORATORIES

5800 Uplander Way

Culver City, CA 90230-6608

Program Director, RDL
Gary Moore

Program Manager, AFOSR
Major David Hart

Program Manager, RDL
Scott Licoscos

Program Administrator, RDL
Gwendolyn Smith

Submitted to:

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

Bolling Air Force Base

Washington, D.C.

December 1995

PREFACE

Reports in this volume are numbered consecutively beginning with number 1. Each report is paginated with the report number followed by consecutive page numbers, e.g., 1-1, 1-2, 1-3; 2-1, 2-2, 2-3.

Due to its length, Volume 5 is bound in three parts, 5A, 5B and 5C. Volume 5A contains #1-23, Volume 5B contains reports #24-44 and 5C contains reports #45-64. The Table of Contents for Volume 5 is included in both parts.

This document is one of a set of 16 volumes describing the 1995 AFOSR Summer Research Program. The following volumes comprise the set:

VOLUME	TITLE
1	Program Management Report
	Summer Faculty Research Program (SFRP) Reports
2A & 2B	Armstrong Laboratory
3A & 3B	Phillips Laboratory
4	Rome Laboratory
5A, 5B & 5C	Wright Laboratory
6A & 6B	Arnold Engineering Development Center, Wilford Hall Medical Center, and Air Logistics Centers
	Graduate Student Research Program (GSRP) Reports
7A & 7B	Armstrong Laboratory
8	Phillips Laboratory
9	Rome Laboratory
10A & 10B	Wright Laboratory
11	Arnold Engineering Development Center, Wilford Hall Medical Center and Air Logistics Centers
	High School Apprenticeship Program (HSAP) Reports
12A & 12B	Armstrong Laboratory
13	Phillips Laboratory
14	Rome Laboratory
15A&15B	Wright Laboratory
16	Arnold Engineering Development Center

SFRP FINAL REPORT TABLE OF CONTENTS i-xiv

SFRP FINAL REPORTS

E-MORPH: A SYSTEM FOR EVOLVING STRUCTURAL FEATURE DETECTORS FOR AUTOMATIC TARGET RECOGNITION

Mateen M. Rizki
Associate Professor
Department of Computer Science and Engineering

College of Engineering and Computer Science
Wright State University
Dayton, Ohio 45435

E-MORPH: A SYSTEM FOR EVOLVING STRUCTURAL FEATURE DETECTORS FOR AUTOMATIC TARGET RECOGNITION

Mateen M. Rizki
Associate Professor
Department of Computer Science and Engineering
Wright State University

Abstract

E-MORPH is multi-phase evolutionary learning system that evolves cooperative sets of features detectors and combines their response using a simple discriminant function to form a complete pattern recognition system. The learning system operates on multi-resolution images that are formed by applying a Gabor wavelet transform to a set of grayscale input images. To begin, candidate target/nontarget chips are extracted from the multi-resolution images to form a training set and a test set. A population of detector sets is randomly initialized to start the evolutionary process. Using a combination of evolutionary programming and genetic algorithms, the feature detectors are enhanced to solve a specific recognition problem. This report describes the implementation of E-MORPH and presents recognition results for a complex problem in medical image analysis. The specific recognition task involves the identification of vertebrae in x-ray images of human spinal columns. This problem is extremely challenging because the individual vertebra exhibit variation in shape, scale, orientation, and contrast. E-MORPH generated several accurate recognition systems to solve this task. The techniques used in E-MORPH are generic and can be readily transitioned to many different problem domains.

E-MORPH: A SYSTEM FOR EVOLVING STRUCTURAL FEATURE DETECTORS FOR AUTOMATIC TARGET RECOGNITION

Mateen M. Rizki
Associate Professor
Department of Computer Science and Engineering
Wright State University

INTRODUCTION

The foundation of a robust pattern recognition system is the set of features used to distinguish among the given patterns. In many problems, the features are predetermined and the task is to build a system to extract the selected features and then classify the resultant measurements. In automatic target recognition problems, the identification of a set of robust, invariant features is complicated because the shape and orientation of the objects of interest are often not known *a priori*. As a result, a human expert is responsible of examining each problem to formulate an effective set of features and then build a system to perform the recognition task. An alternative to this labor intensive approach of building recognition systems has emerged in the past ten years that uses learning algorithms such as neural networks and genetic algorithms to automate the process of feature extraction. There are many advantages to the automated construction of recognition systems over techniques that rely solely on human expertise. Automated approaches are not problem specific. Consequently, once an automated system is developed it can be readily applied to similar problems greatly reducing the time needed to solve new recognition problems. Automated systems are capable of producing solutions that are comparable to the customized solutions created by human experts, but the solutions formed by these systems are often non-intuitive and quite different from the solutions formed by human experts. In many applications, this is a drawback because it is not possible to describe how the solution is obtained. This is also a strength of the automated approach. Automated techniques are unbiased. The features selected to solve problems represent alternative designs based on the structural and statistical attributes of the data. The fact that different features are selected suggests that automated systems are capable of exploring different regions of the space of potential solutions.

E-MORPH [Rizki et. al. 1993, 1994] is an evolutionary learning system that generates pattern recognition systems using several different learning paradigms to automatically perform feature extraction and classification. A robust set of features is identified using a population of pattern recognition systems. Each system is composed of a collection of cooperative feature detectors and a classifier that evolves under the control of a user provided performance measure. The performance measure is typically tied to recognition accuracy, but additional constraints may be included such as complexity measures to sculpt specific types of solutions. The recognition systems compete for survival based on their performance. Successful systems have a higher probability of survival and contribute more information to future generations. The structural and statistical information gathered by each recognition system during the evolutionary process is passed to the next generation through a process of reproduction with variation. The most successful recognition systems are combined to form new recognition systems that are often superior to either parental unit. Two opposing forces operate in the evolutionary process: exploration and exploitation. By recombining successful solutions during reproduction , each generation contains recognition systems that are more capable of exploiting the performance

measure and solving the recognition task. The reproductive process is imperfect, variations in the new recognition systems are created by mutating the structure of the feature detectors. Each new recognition system contains pieces of past successful designs with variations that explore alternative designs. The process of reproduction with variation and selection continues until the best recognition system in the population achieves a satisfactory level of performance.

E-MORPH LEARNING SYSTEM

The overall design of our pattern recognition system is shown in Figure 1. Grayscale images pass through a Gabor wavelet [Gabor, 1946] transformation module that explodes the image into parallel streams of registered images. The Gabor wavelets are oriented bandpass filters that represent an optimal compromise between positional and spatial frequency localization. The Gabor module effectively organizes the spatial frequency and positional information present in the raw grayscale imagery into a registered stack of Gabored images. The target detection module then extracts regions from the full Gabor stack to form registered stacks of small chip-images. The pixel values of each chip are then scaled to fall in the range of -128 to +128. Finally, each stack of chips is processed by the target recognition module that extracts features and assigns a label to each chip.

A recognition system is composed of an ordered set of feature detectors and a discriminant function to separate target and nontarget stacks of chips. Each feature detector is represented by a convolution template large enough to cover the area of a chip. The convolution kernel contains a collection of probes points restricted to the values of +1 and -1. The use of a two-valued template allows detectors to explore both geometrical structure and contrast variation. For example, when a positive point (+1) is placed over a bright area of a chip and a negative point is simultaneously located over a dark area of the chip, the convolution operator produces a large output that signifies that the geometry and contrast variation embodied in the template exists at a specific displacement from the center of the chip. By adjusting the positions and values of the probe points, complex structural relationships can be readily identified. The set of detectors
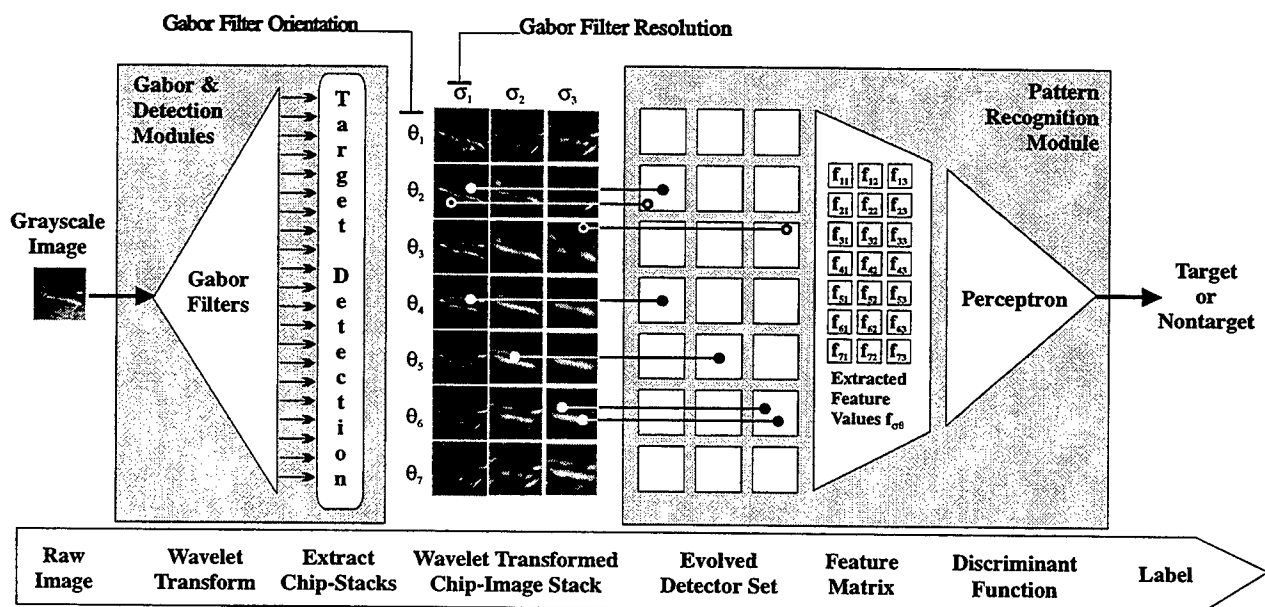

Figure 1 The representation of a pattern recognition system.

present in a single recognition system forms a registered set of convolution templates that serves as a 3D probe. The 3D probe spans multiple registered chips in the wavelet transformed stack and allows the detector set to explore relationships within a single stack-plane or across several stack-planes. For example, since the wavelet transform produces stack-planes corresponding to different spatial frequencies, the 3D probe can exploit multiple resolution levels to identify the presence of a fragmented edge that is not easily recognized at a high spatial frequency but is quite prominent at a lower spatial frequency. When the full set of detectors is convolved with a stack of images, an integer value feature vector is produced. Each detector contributes one value to the vector. By repeating this process for all the image stacks, a feature matrix is created that is passed to the discriminant function.

The E-MORPH system provides candidate detector sets to recognition system module and collects the corresponding system error as shown in Figure 2. The error vector is used to assess the accuracy of the candidate detector set and calculate a performance measure. Performance is defined using a combination of target and nontarget recognition accuracy (see Equation 1).
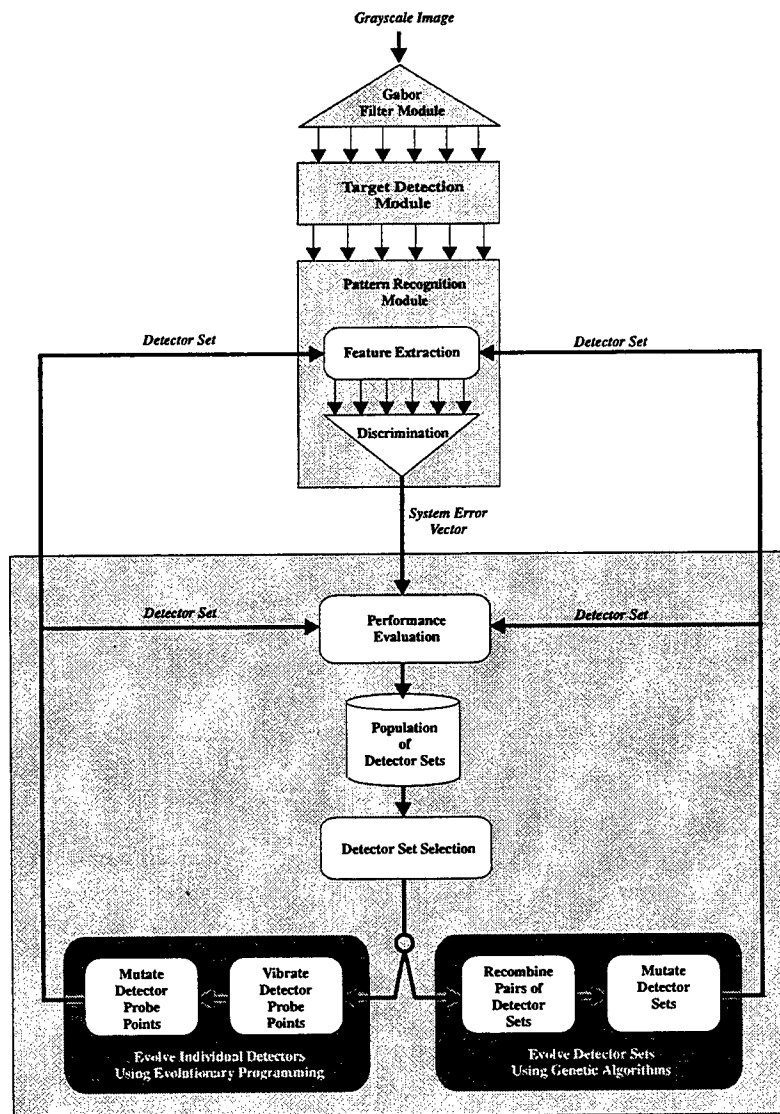

Figure 2 Overview of the E-MORPH learning module.

$$pm_i = \left(\alpha \cdot \frac{t_i}{T} + (1-\alpha) \cdot \frac{nt_i}{NT} \right)$$

Equation 1

The variable t_i is the number of targets correctly classified by the ith recognition system, T is the total number of targets, nt_i is the number of nontargets correctly classified, NT is the total number of nontargets, and α is a weight ($0 < \alpha < 1$).

After all the recognition systems are assigned a performance measure, each recognition system competes for survival with other members of the population. The competition is organized as a tournament that ranks each recognition system based on its performance relative to the performance of other systems in the population. The size of the tournament changes throughout the evolutionary process and is based on the average performance of the population as shown in Equation 2

$$NC = max \left(1, \; M \cdot \frac{\sum_{i=1}^{N} pm_i}{N} \right)$$

Equation 2

In this equation, NC is the number of competitors in each tournament, N is the population size, and M is a user imposed upper limit on the number of competitions ($M <= N$). Each recognition system must win as many conflicts as possible to increase its chance for survival. The number of competitions won or lost is calculated using equation 3. In these

$$win_i = \sum_{k=1}^{NC} \left[\left(\frac{pm_i}{pm_i + pm_{2 \cdot N \cdot U(0,1)}} \right) < U(0,1) \right]$$

Equation 3

local competitions, the chance of winning is proportional to the ratio of the performance measures of the recognition system and its competitor. For example, if a recognition system's performance (pm_i) is high and a randomly selected competitor's performance ($pm_{2*N*U(0,1)}$) is low, then the probability that the ratio is greater than a value drawn from a uniformly distributed random variable $U(0,1)$ is also high. When the relation is satisfied, the recognition system wins the pairwise competition. Limiting the tournaments to a subset of the population reduces the possibility of premature convergence of the evolutionary process. When the average performance of the population is poor, the number of individuals in each tournament is small and a marginally better recognition system does not have the opportunity to dominate the population. The pairwise competition used within each tournament tends to maintain a diverse population of recognition systems because marginal individuals always have a small probability of survival. The final selection for survival is based on a ranking of the number of conflicts won by each recognition system. The sets with the greatest number of victories survive to the next learning cycle.

E-MORPH uses two different techniques to alter the structure of the detector set contained in each recognition system. The position of the probe points in the convolution templates within a detector set are varied using evolutionary programming [Fogel, 1991], and the collection of convolution templates that forms a detector set is varied using a genetic algorithm [Holland, 1975]. These techniques are combined to explore different aspects of detector population. The evolutionary programming algorithm begins by cloning each member of the surviving population. The structure of the individual detectors within each clonal set is manipulated by vibrating the position of each probe point as shown in Figure 3. The extent of the vibration of each probe point is modulated by a Gaussian envelope.

A complete E-MORPH learning cycle consists of several sub-cycles of the detector optimization using the evolutionary programming algorithm (EP) followed by several sub-cycles of detector set recombination using the genetic algorithm (GA) (see Figure 2). The purpose of the EP algorithm is to systematically improve the position, type, and number of probe points in the active convolution templates. This is accomplished using a controlled vibration of the position of
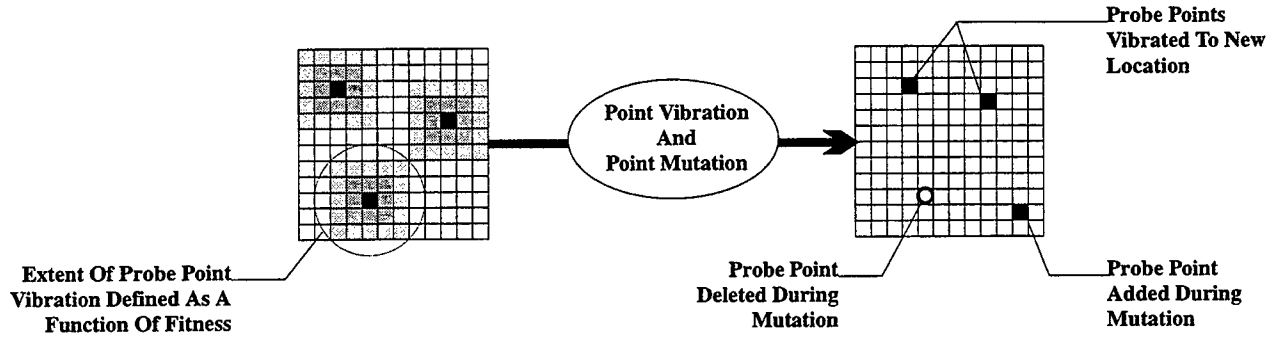
Figure 3 Using Evolutionary Programming to evolve a convolution kernel.

the points in each template followed by a series of random mutations that add and/or delete points (see Figure 3). To begin an EP cycle, the existing population of N detector sets is reproduced to form an extended population of 2N detector sets. Each detector set in the extended population is subjected to random variations. The amount of variation is inversely proportional to the performance of the parental detector set and controlled by Equation 4.

$$x_{j,k} = max\left(min\left(x_{j,k} + \left(\frac{X_{size_j}}{2} \cdot (1 - pm_i) \right) \cdot N(0,1), \ X_{max_j} \right), X_{min_j} \right)$$

$$y_{j,k} = max\left(min\left(y_{j,k} + \left(\frac{Y_{size_j}}{2} \cdot (1 - pm_i) \right) \cdot N(0,1), \ Y_{max_j} \right), Y_{min_j} \right)$$

Equation 4

The value $(x_{j,k}, y_{j,k})$ is the position of the kth point in the jth template, (X_{size}, Y_{size}) is size of the template, (X_{min}, Y_{min}) is the location of the lower left corner of the template, (X_{max}, Y_{max}) is the location of the upper right corner of the template, $(1 - pm_i)$ is the complement of the performance measure of the ith detector set, and $N(0,1)$ is a normally distributed random variable with a mean of zero and a variance of one. To update a probe point's position, the mean of the random variable is set to the value of the initial position of the probe point and the variance is scaled to fall into the range from zero to one-half the template size. Using this technique, when the performance measure is low, the potential extent of variation is high. The potential for variation is reduced as the performance increases. If the performance reaches one, the potential for variation is zero and the template's point configuration is frozen. This approach to adjusting the structure of a template is similar to the process of simulated annealing where gradual improvements in the population performance shutdown the process of random variation as a solution is formed.

The vibration process is only capable of adjusting the position of existing probe points. The second step of the EP phase is mutation that adds and/or deletes probe points to alter the complexity of the templates. Point mutation occurs immediately after the template points are vibrated. The amount of mutation is controlled by Equations 5 and 6.

$$p_a = max\left(\beta_a, \ \mu_a \cdot (1 - pm_i)^{\sigma_a} \right)$$

Equation 5

$$p_d = max\left(\beta_d, \ \mu_d \cdot pm_i^{\sigma_d} \right)$$

Equation 6

The probability of a point mutation (p_a, p_d) is calculated using a user defined multiplier (μ_a, μ_d) that provides an upper limit on the mutation rate and a scale factor (σ_a, σ_d) that shapes the probability curve. In addition, the user must supply

a baseline value (β_a, β_d) that sets a lower limit on the amount of mutation. For example, if the user sets β_a=0.2, μ_a=0.8, and σ_a=1.0, then as the performance (pm_i) of the detector set rises, the mutation rate will start at 0.8 and fall off linearly to 0.2. By adjusting the value of σ_a, the rate of change of the mutation probability can be altered to remain high or low for larger ranges of performance. The mutation rate parameters must be set to correspond to the initial conditions used to form the population of detectors. If the detector set is initialized with a limited number of probe points, the probability of addition should be larger than the probability of deletion. This will bias the mutation rate toward addition and cause the detectors to grow in complexity.

After all of the templates in a detector set are vibrated and subjected to mutation, the set is placed in the recognition module and a Perceptron [Minsky and Papert, 1988] is trained to form a linear discriminant that separates target and nontarget chip-images. The error vector returned by the recognition module is used to assign a performance measure to the modified detector set. This process is repeated for each detector set in the extended population. Finally, the N parental detector sets compete with the N offspring detector sets in a tournament for survival. The top ranked N detectors are preserved and the evolutionary programming cycle begins again. When the EP phase of E-MORPH terminates, each member of the population consists of a cooperative set of convolution templates.

The GA phase of E-MORPH exchanges convolution templates between pairs of detector sets to accelerate the learning process. To begin, a selection vector is formed to simulate a biased roulette wheel sampling process. The number of times a detector set appears in the selection vector is determined by the ratio of the set's performance measure to the average performance of the population. For example, if the the a detector set's performance is 0.8 and the average is 0.4, then the detector appears in the vector twice (0.8/0.4=2). If the ratio does not yield a whole number (e.g. 0.8/0.5=1+0.6), the remaining fraction is compared to a uniformly distributed random variable. If the fraction is larger, an additional copy of the detector is added to the selection vector. Performance is scaled prior to forming the selection vector to prevent a marginally better detector set from dominating the selection process [Goldberg,1989]. Reproduction begins by drawing random pairs of parental detector sets from the selection vector. The set of convolution templates in each parent are analogous to a biological chromosome, and the individual templates are similar to genes. The uniform crossover operation consists of stepping through the detector sets of both parents, producing a copy of the templates at the corresponding positions, and assigning the copies to a pair of offspring detector sets at random. This is illustrated in Figure 4. The color coded (shades of gray) set of parental convolution templates is shown to the left and the mixture of templates duplicated in the offspring are shown to the right. The convolution templates are ordered to correspond to fixed template positions in the Gabored stack of chip images. When the crossover operation exchanges portions of the detector set, the potential exists to bring meaningful pieces of two complex geometrical structures together.

After the recombination process, each offspring detector set is mutated by adding probe points to templates that are empty or by deleting all the probe points in a template. The EP phase is limited to vibrating and mutating convolution templates that are active (contain at least one probe point). The GA phase is responsible for activating and/or deactivating templates without altering the internal point distribution of active templates. The amount of addition and deletion is controlled by Equations 7 and 8 respectively. These equations are similar to the distributions used to control point

$$p_A = max\left(\beta_A, \mu_A \cdot \left(1 - \overline{PM}\right)^{(1+(dt_i - \overline{DT})/\sigma_{DT})}\right) \qquad \textbf{Equation 7}$$

$$p_D = max\left(\beta_D, \mu_D \cdot \overline{PM}^{(1+(dt_i - \overline{DT})/\sigma_{DT})}\right) \qquad \textbf{Equation 8}$$
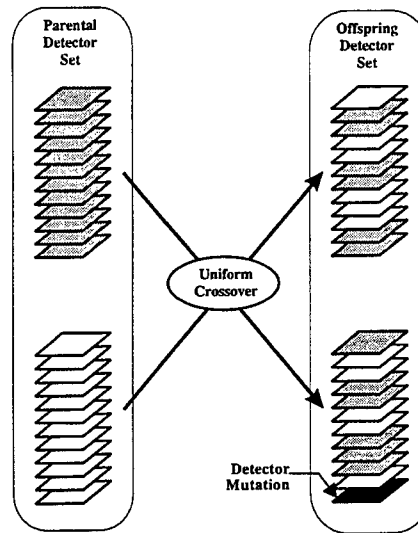
Figure 4 Detector Set Variation Using Genetic Algorithms.

mutation during the the EP phase. A user supplied baseline mutation rate (β_A, β_D) insures that some level of variation continues throughout the learning process. The upper limit on the mutation rate (μ_A, μ_D) is used to avoid introducing an excessive amount of variation during the early stages of the evolutionary process before the performance begins to increase. The probability of activating or deactivating a template within a selected detector set is controlled by the average performance of the detector set population ($\overline{PM}$), the average number of activate templates ($\overline{DT}$) in the population, the variance (σ_{DT}) in number of templates in the population, and the number of active templates in the individual detector set (dt_i). Equations 7 and 8 scale the mutation rates to fall in the interval [β, μ]. The complexity term modulates the mutation rate by lowering it when an individual detector set's number of activate templates exceeds the population average and raising it when the number is below average. These equations work to balance the opposing forces of performance and complexity by accelerating or decelerating the rate of variation of the detector sets.

The GA phase begins with N detector sets and combines N/2 pairs of parental set to form an extend population of 2N sets. Each member of the extended population is evaluated using the same procedure described for the EP phase. A tournament selection process is applied to rank the entire population and the N top-ranked detector sets are preserved for the next cycle of the GA algorithm. When the GA phase is complete, each detector set consists of combinations of templates that proved useful in the recognition process. In addition, the average number of active templates in detector set population will have evolved to produce detector sets with higher recognition accuracy.

The user can set parameters to control the number of EP and GA sub-cycles that occur within each E-MORPH learning cycle. There is a tradeoff between the EP and GA phases. The user can increase the sensitivity of the individual templates by increasing the number of passes through the EP phase relative to the number of passes through the GA phase. Alternatively, the user may elect to spend more computational resources adjusting the average complexity of the detector sets by increasing the number of passes through the GA phase. It is difficult to select an appropriate mixture of passes because the evolutionary learning process is dynamic. During the early stages of evolution, it is not likely that the average number of active templates in the population is suitable for the recognition task. If the user arbitrarily increases the number of EP passes, the probe point density will increase to compensate for the lack of active templates. This will produce customized solutions that tend to perform well on training sets and poorly on test sets. If the number of GA

cycles is too large, the average number of templates per detector will increase to compensate for the inadequate distribution of probe points within each template. Our solution to this problem is to use a large number of E-MORPH learning cycles and a small but equal number of passes through the EP and GA within each learning cycle. A better solution would be to implement an adaptive control mechanism that evaluates the relative contribution of each phase throughout the evolutionary process and dynamically adjusts the length of each phase.

EXPERIMENTAL DESIGN

To demonstrate our technique for generating a pattern recognition system, a target recognition task in medical imagery is presented. Specifically, the task is to locate and measure the deterioration in a patient's spinal column using radiographic images. This problem is difficult because the images contain large variations in contrast and the target vertebrae are extremely distorted. Examine the sample images shown in Figure 5. The vertebrae located in the image on the right are slightly larger than the vertebrae in the image on the left. In addition, there are 3D effects caused by the tilting of the vertebrae as evidenced by the small elliptical shape visible on some vertebrae. To simplify this recognition task, we restricted our attention to the problem of locating the vertebrae. In particular, our goal is to detect the edges that define the gaps between vertebrae.

There were 48 variable size images used in this experiment. To begin, a central region measuring 512x768 pixels was extracted from each image. Each image was then processed using a set of two-dimensional self-similar Gabor functions. The Gabor functions are defined as the product of a radially symmetric Gaussian function and a complex sinusoidal wave as shown in Equation 9.
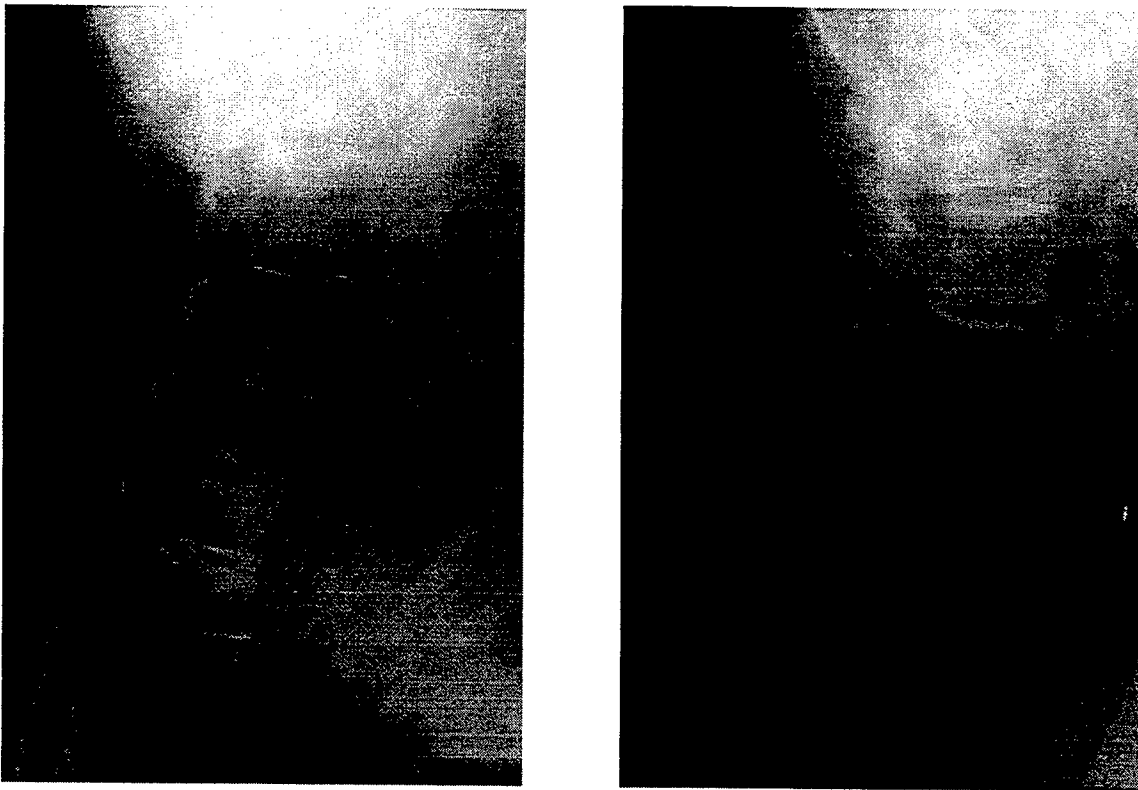

Figure 5 Sample x-ray images used in the target recognition experiment.

$$G(x, y; m, n) = e^{-r^2/s^2} \cdot e^{ikr}$$

$$r = [x, y] \qquad k = \frac{1}{\lambda} \cdot \left[cos\left(\frac{(3+m)\pi}{12} \right), sin\left(\frac{(3+m)\pi}{12} \right) \right] \qquad \text{Equation 9}$$

$$m = 0...6 \qquad n = 1...3 \qquad \lambda = 2 \cdot S = 4 \cdot n$$

The vector r represents a displacement vector and k is a propagation vector. The wavelength λ is specified in pixel units and is related to the width of the Gaussian envelope. The self-similar property of the Gabor function is evident in the resolution levels that are scaled versions of each other. In this experiment, we utilized seven orientations starting at 45° and continuing through 135° in increments of 15°. The resolution levels were set to λ equals 4, 8, and 12 pixels. The choice of parameters correspond to the general structure of the vertebrae in the grayscale images, but no attempt was made to optimize system performance by adjusting these values. Using these parameters, we generated 21 Gabor filters and applied them to the 512x768 grayscale images.

The Gabor output images are formed by convolving each Gabor filter with an input image. This is accomplished by computing the product of the Fourier transform of the Gabor filters with the Fourier transform of the image and taking the inverse Fourier transform of the product. This results in complex image consisting of real and imaginary values at each pixel location.

The real valued part of complex image corresponds to the cosine term of the propagation vector and the imaginary part is associated with the sine term. The cosine term produces a wave form that exhibits reflexive symmetry across the center of filter in the direction of the wave propagation while the sine term yields an asymmetric wave form. Both the sine and cosine act as edge detectors, but only the sine version is sensitive to the sign of the edge gradient. In this experiment, we use the magnitude of the complex image which produces a smeared version of the sine and cosine edge images due to the phase difference between these trigonometric functions. To complete the preprocessing, each of the resultant real valued magnitude images was byte scaled to map the dynamic range of the intensity into 256 discrete levels of gray.

The automated target detection module is not implemented in this version of E-MORPH. Targets were located by examining individual raw grayscale image to identify the approximate center of each vertebrae gap. Chip images were then formed by defining a 128x128 bounding box around each target center. This process was repeated for each input image to produce 230 target chips. One set of nontarget chips was formed by displacing each target chip in some random direction by a minimum distance of 16 pixels and a maximum distance of 48 pixels. A second set of nontarget chips was formed by extracting 230 randomly located chips from the input images. When the three categories of chips (targets, displaced nontargets, and randomly selected nontargets) were combined, the result was a set of 690 images.

A training set consisting of 345 chips was formed by selecting 115 images at random from each of the three categories as shown in Figure 6. The remaining 345 chips were placed in a test set (see Figure 7). The grayscale version of the training and test sets are shown to aid the reader. E-MORPH does not operate on the raw grayscale images. The actual set of training images consisted of the 345 chip-stacks that were extracted from the corresponding positions in the Gabor filtered images. As a result, the actual number of individual chips in both the training and test sets was 21 * 345 = 7245 chip images. A few sample Gabor stacks corresponding to the three categories of images (targets, displaced nontargets and random nontargets) are shown in Figure 8. Notice the amount of variation between the two sample target images
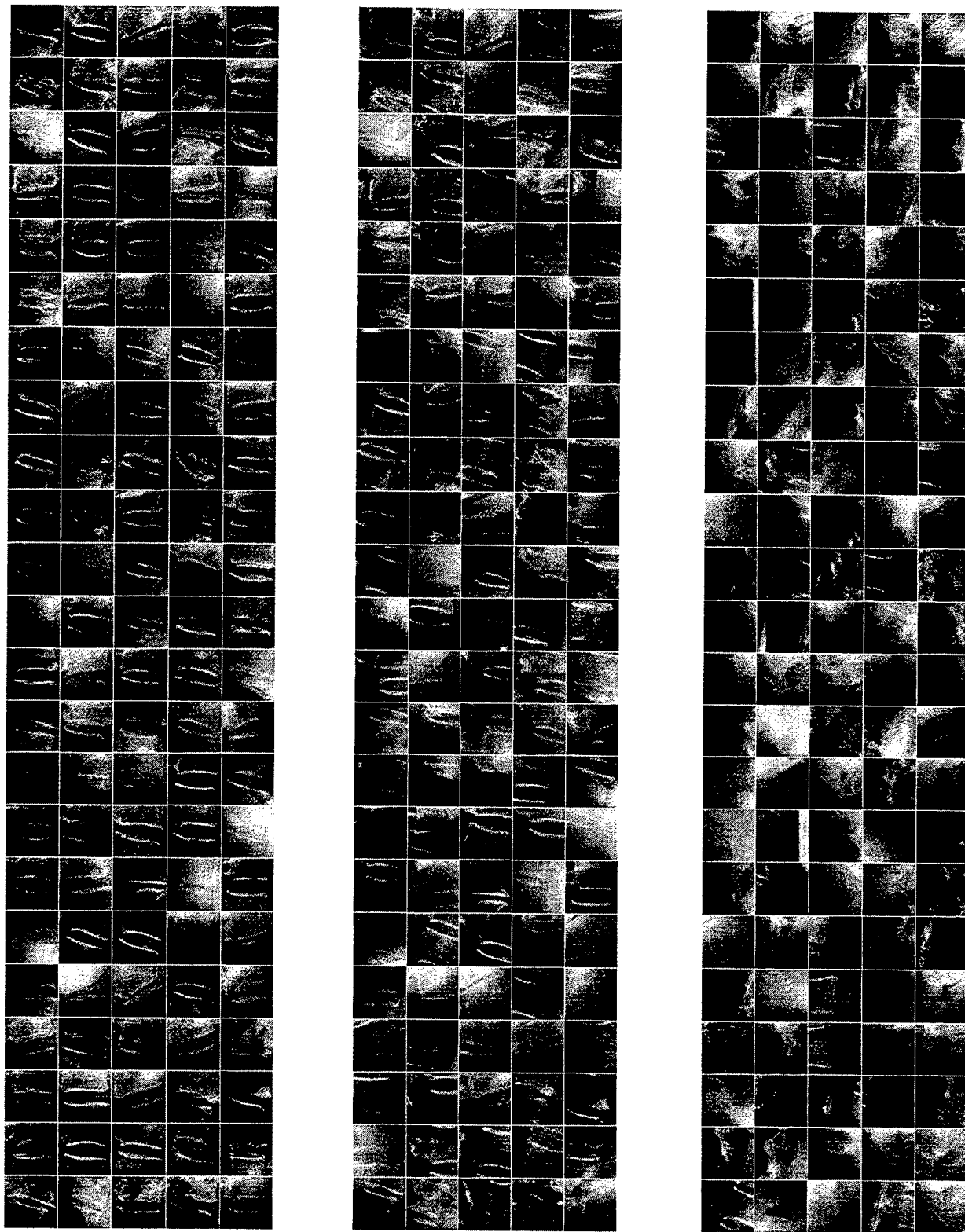
The training set composed of 345 grayscale chip image. The training set consists of 115 target chips
Figure 6 (left column), 115 displaced nontarget chips (middle column), and 115 randomly selected nontarget chips
(right column). Each chip is 128x128 pixels.

The test set composed of 345 grayscale chip image. The test set consists of 115 target chips (left column),
Figure 7 115 displaced nontarget chips (middle column), and 115 randomly selected nontarget chips (right column).
Each chip is 128x128 pixels.

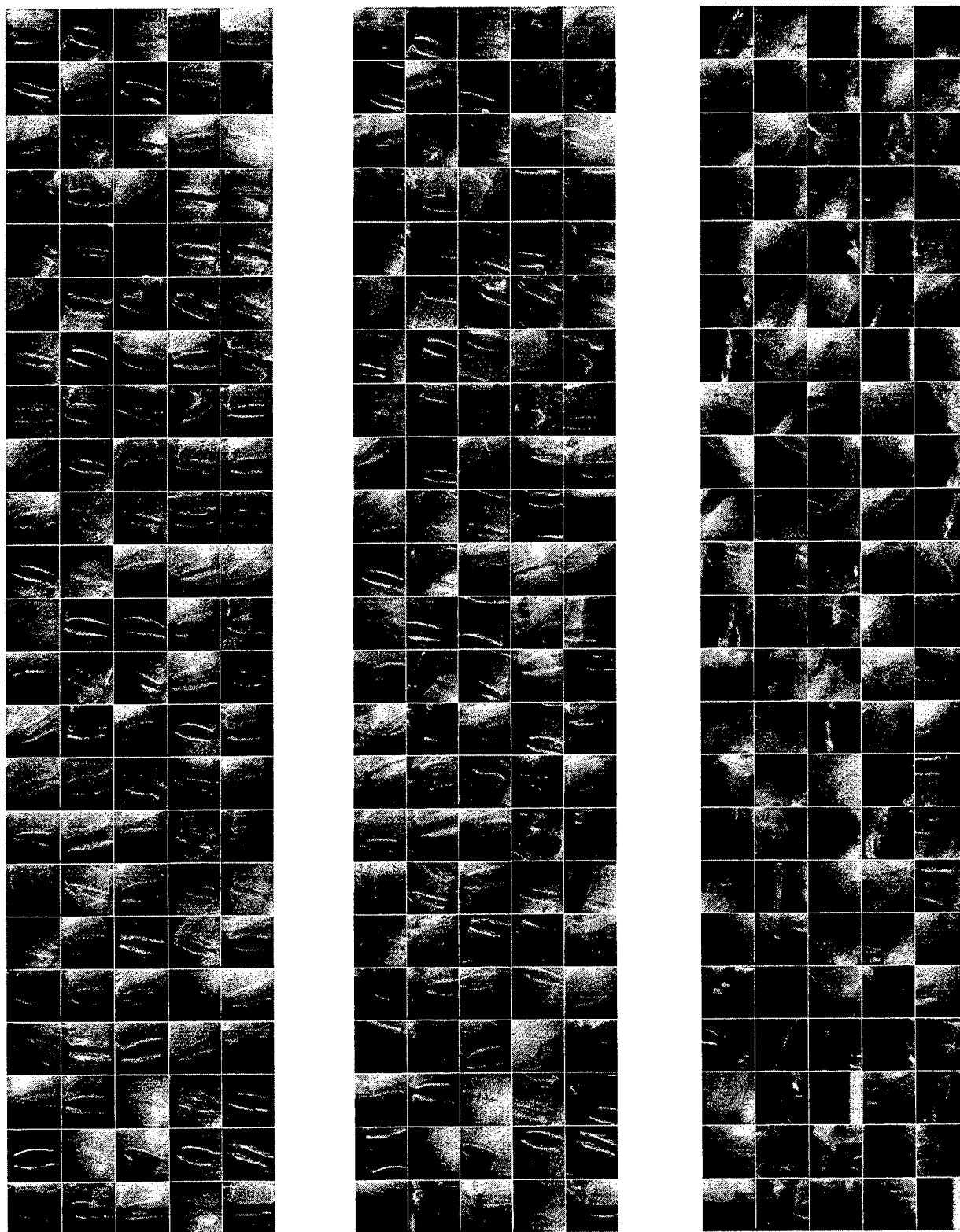shown at the top of the figure. Also observe the amount of similarity between a target image and its displaced nontarget counterpart shown at the left of the figure. The variation among the target images can easily exceed the variation between targets and displaced nontargets, making it extremely difficult to discriminate between target and nontarget images.
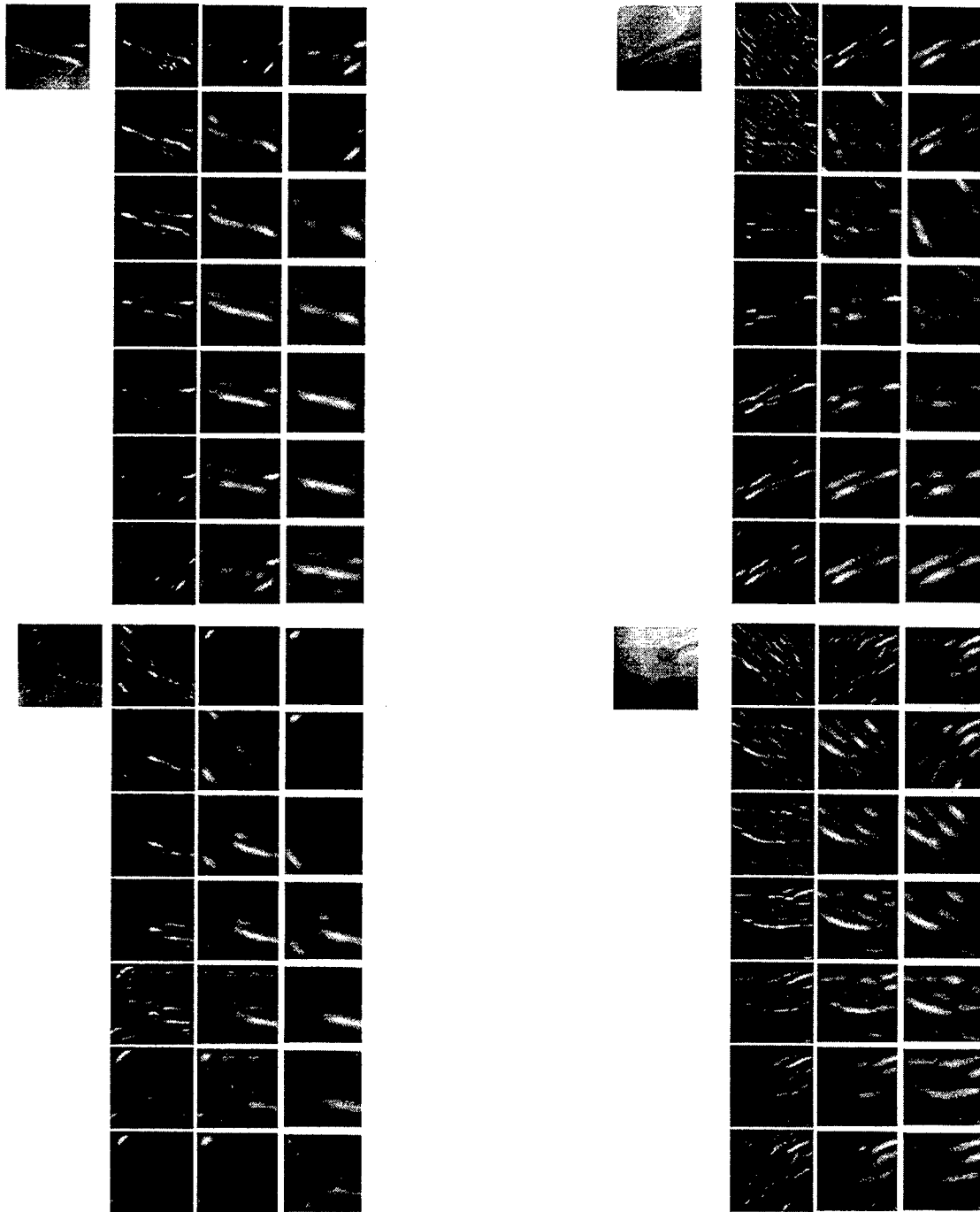
Sample Gabored chip-images. Four sample stacks of images are shown. The small image to the left of each block of Gabored chips is the corresponding grayscale chip. The columns of each block of Gabor images
Figure 8 represent different resolutions and the row are different orientations (45°-135°). The top pair of blocks are targets, the bottom left block is a displaced nontarget corresponding to the target image directly above it, and the bottom right block is a random nontarget.

To begin the learning experiment, a population of 32 detector sets was generated. Each set was initialized by placing a probe point at a random location in three randomly selected templates. Theses templates were then evaluated by convolving them with the Gabor training chips to generate feature vectors that were passed to the Perceptron to compute a recognition accuracy for each detector set. The performance measure was computed by giving equal weight to target and nontarget accuracy .

An E-MORPH learning cycle consisted of five EP sub-cycles followed by five GA sub-cycles. The minimum and maximum values for the mutation rates in the EP phase were set to 0.2 and 0.6 for point addition and 0.05 and 0.2 for point deletion. The mutation rate limits for the GA phase were set to 0.2 and 0.6 for detector activation and 0.05 and 0.1 for detector deactivation. Each pass through a sub-cycle produced 32 mutated detector sets creating an extended population (parents and offspring) of 64 detector sets. The extended population was ranked using tournament selection and the top 32 detectors were saved to start the next sub-cycle. At the end of each pass through a sub-cycle, performance statistics for the population were saved. The experiment was run for a total of 22 learning cycles (110 EP sub-cycle and 110 GA sub-cycles).

The average recognition accuracy for the population produced during the evolutionary learning process is shown in Figure 9. The performance is displayed on even numbered learning cycles. Initially the average nontarget recognition was approximately 75%, but the target recognition accuracy was less than 10% (not shown on the graph). This high nontarget recognition accuracy is an artifact. Initially, most detectors produced a negative (< 0) response to almost every chip, and a negative response on a nontarget was recorded as a correct answer even though the detector had no ability to discriminate between targets and nontargets. After 20 sub-cycles, target recognition accuracy began to improve. Training accuracies were approximately 77% for targets and 84% for nontargets. The test set recognition accuracies lagged behind at 68% for targets and 76% for nontargets. As the evolutionary process continued, average recognition accuracies slowly increased to their final values of 87% / 92% (target / nontarget) accuracy on the training sets and 78% / 88% (target / nontarget) accuracy on the test set. The difference between the training set and test set
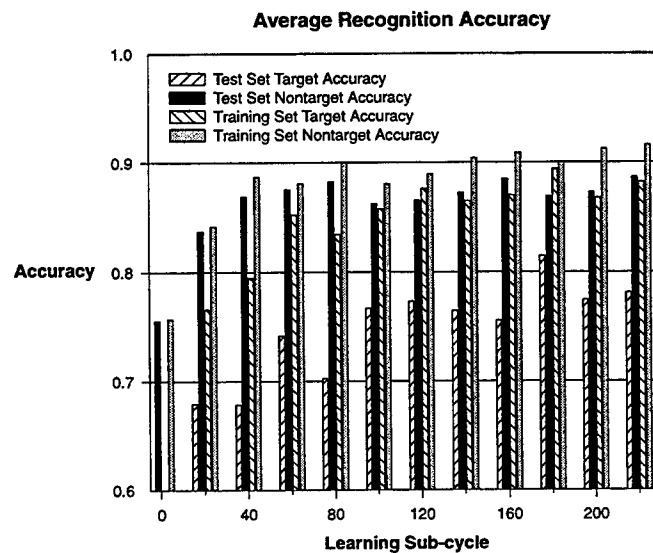

Average Recognition Accuracy

Figure 9 Average recognition accuracy.

performance suggests that the detector sets are not generalizing to the test set. The difference between the training and test set target accuracy is the most obvious problem. The explanation of this difference is obvious if we review the response vectors for the individual detector sets at the end of the learning process (see Figure 10). The majority of errors occur among the target and displaced nontarget portions of the the training and test sets. To resolve these errors during the training process, the detector sets have to become highly customized. This specialization tends to limit the ability of the detectors to generalize when they are applied to the test set.

The best combined recognition accuracy achieved at the end of the experiment was 93% on the training set and 86% on the test set (see Figure 11). The fluctuations in the performance shown in this figure are due to the tournament selection process. The top performer in one generation can be eliminated from the population during the competition for survival. Keeping the most accurate solution is one way to eliminate this problem, but notice the detector set that produces the best performance on the training set appears in sub-cycle 160 and has a rather marginal test set accuracy. Again this problem is related to the conflict between specialization and generalization caused by the similarity of the target and displaced nontarget images. One simple approach to deal with this problem is take a majority vote using the top three or five detectors. We tested this approach, and it increased recognition accuracy by 3-5%.
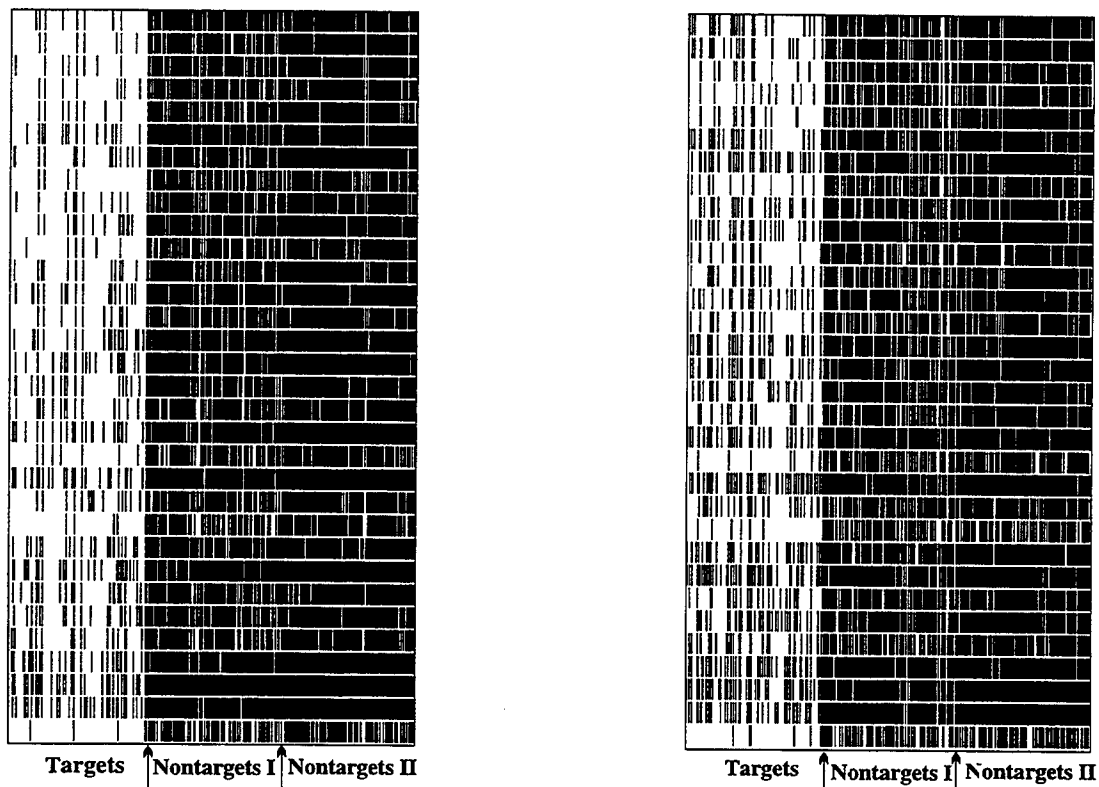

Targets | Nontargets I | Nontargets II   Targets | Nontargets I | Nontargets II

Figure 10 **Final response vectors.** The response vectors for the final population of 32 detector sets are shown. The training set response is shown in the left box and the corresponding test set response is shown in the right box. Responses are rank ordered by training performance. Each row represents the response of the detector set to 345 chip images grouped into targets, nontargets I (displaced targets), and nontargets II (random targets). A thin black stripe in the target area or thin white stripe in the nontarget area represents one error.
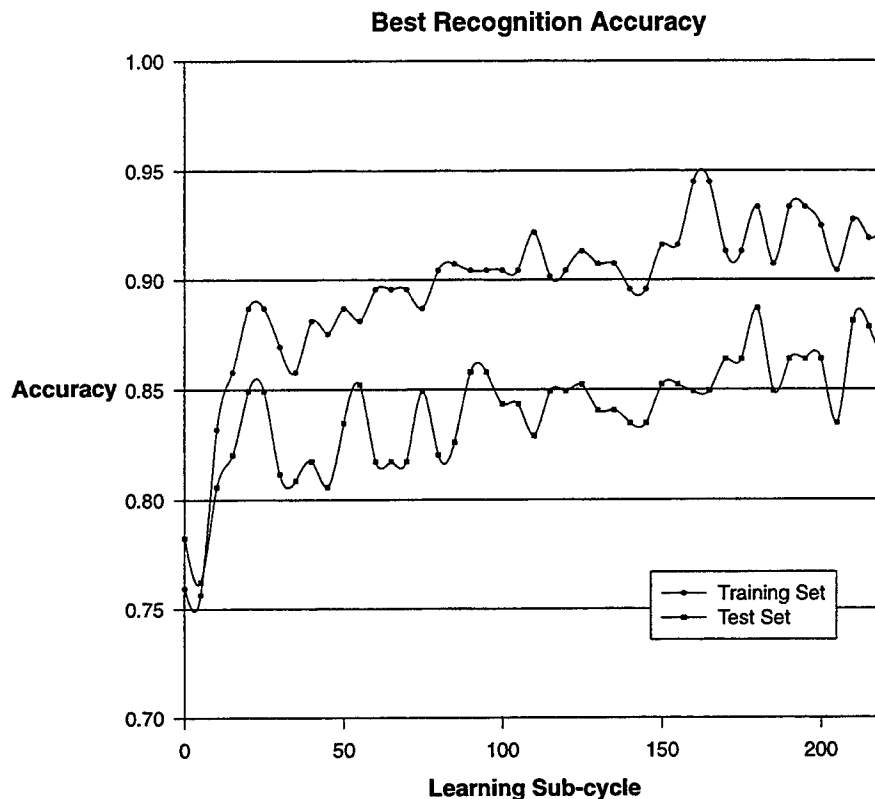
Best Recognition Accuracy

Figure 11 The recognition accuracy of the best detector set.

The structure of a few detector sets from the final population are shown in Figures 12 and 13. In Figure 12, three different detectors are superimposed on the first image in the training set. Notice the common substructure present in these three sets. This is to be expected at the end of the experiment as the system converges to a solution. What is surprising is the amount of subtle variation that still exists. For example, examine the rightmost template in the second to the last row of each detector set. These templates have the same basic footprint, but they are not identical. The basic pattern embodied in this template appears to align nicely with the vertebra gap, but it is difficult to predict whether this is a useful template by examining a single Gabor image.

In Figure 13, one template is superimposed on two target images and a displaced nontarget image. This is the same detector set that appears in the upper left corner of Figure 12. Again examine the rightmost template in the next to the last row. Notice how the template's probe points align with the vertebrae edges in both targets but do not align with the displaced nontarget image. It is difficult to draw a general conclusion, but the distribution of points among different detector sets suggest that the cluster of points are sensing a on-off-on type of relationship embodied in the edge of a vertebrae. Occasionally, a cloud of points of the same type appear in a template as seen in the fifth row of the leftmost column. These clouds appear to sense large active or inactive areas in the chip images.

Although the points are scattered throughout the template. In general, a greater number of points appear in the lower resolution templates where edges are more pronounced. Also clouds of points are more common in the lower resolution templates. This suggests that the high frequency Gabor images are too sensitive to variations among the training images
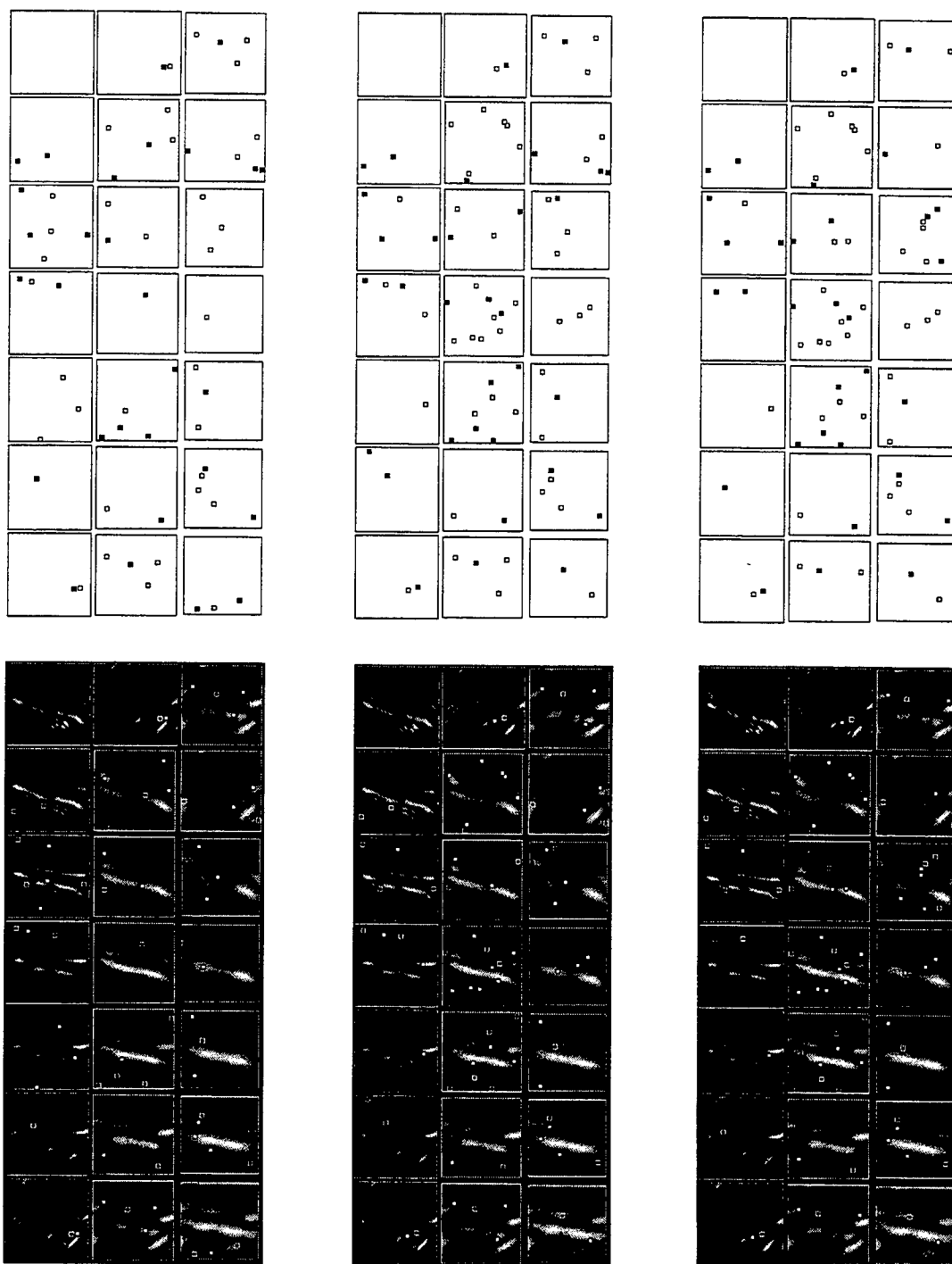
45-17

Figure 12 An E-MORPH generated detector sets applied to a target image. These three detector sets (top row) are taken from the population at the end of the final learning cycle. Each template probe point is shown superimposed on the corresponding Gabor target chip (bottom row). The dark center-white surround is a -1 point and the white center-dark surround is a +1 point.
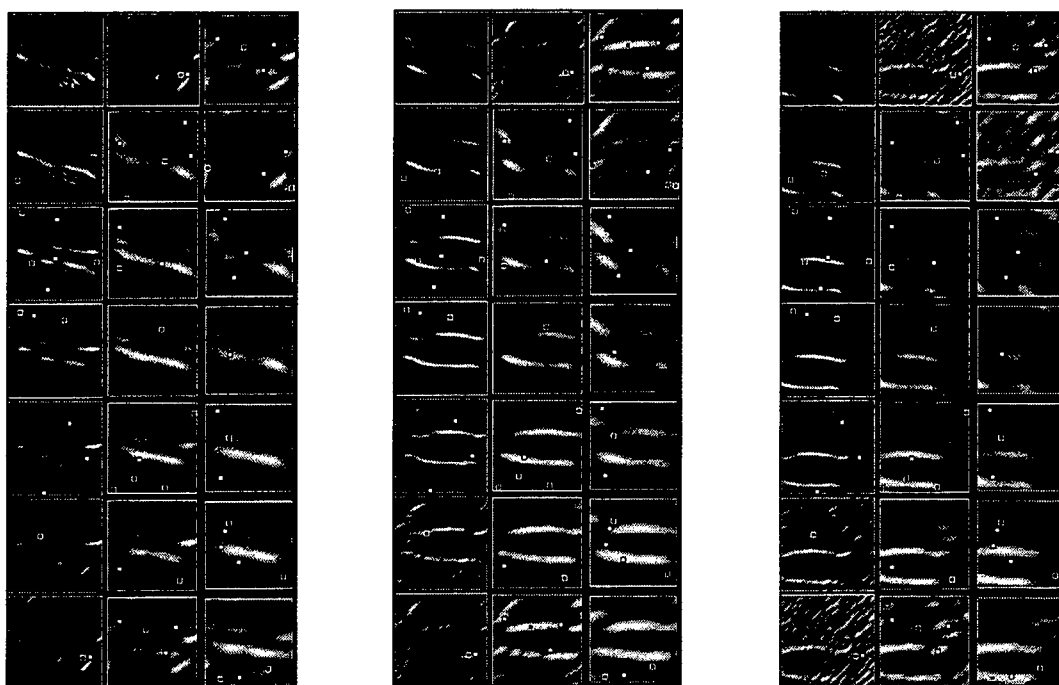
Figure 13 An E-MORPH generated detector sets applied to three different Gabor images. A single detector set is shown applied to two target images (left and middle) and a displaced nontarget image (right).

to be useful in the general solution, but they may be very useful in separating a target from its corresponding displaced nontarget image.

DISCUSSION

E-MORPH successfully generated a pattern recognition system capable of solving an extremely difficult problem in medical imaging. In particular, x-ray images of human spinal columns are processed to locate vertebrae using Gabor filters to form a multi-resolution edge image. E-MORPH was then used to select features from the Gabor images and assemble pattern recognition systems. The learning process starts with a random assemblage of convolution templates that are enhanced using a hybrid evolutionary learning algorithm that exploits the strengths of both evolutionary programming and genetic algorithms. At the end of the experiment, the evolved population of feature detectors includes a detector that produces a 93% recognition accuracy on the training set and 86% accuracy on the independent test set. There are other members of the population that produce similar results. The performance can be improved by combining the results of several detector sets in a voting process, but a better solution is to add a more sophisticated classifier to deal with the in-class variability of the data set. Clearly, the capability of the linear discriminant used to separate targets and nontargets is limited and forces E-MORPH to compensate by generating detector sets customized to the training data.

The overall structure of the detectors generated by E-MORPH appears to correspond to both the geometrical and contrast variations present in the images, but the complexity of the training set makes it difficult to analyze the behavior of the

individual detector sets. We believe the detector sets are using a complex combination of geometric structure, contrast variation, and statistical averaging of the lower spatial frequencies present in specific Gabored images to guide the search process.

There is no single approach that solves all problems in automatic target recognition. E-MORPH represents one viable alternative. Solutions generated using our evolutionary learning algorithm are quite different than solutions produced by human experts. This suggests that human experts may not be using all of the available information to develop robust pattern recognition systems. In future work, we hope to explore the possibility of combining human expertise with the evolutionary search process to access these design alternatives. This hybrid approach to design may ultimately produce recognition systems with performance superior to any in use today.

ACKNOWLEDGMENT

REFERENCES

Fogel, D. B. (1991). *System Identification Through Simulated Evolution: A Machine Learning Approach to Modeling*, Needham, MA: Ginn Press.

Gabor, D. (1946). "Theory of Communication", J.I.E.E., 93:429-459.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*, Reading, MA: Addison-Wesley.

Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems*, Ann Arbor, MI: The University of Michigan Press.

Minsky, M. L. and S. A. Papert (1988). *Perceptrons*, Cambridge, MA: The MIT Press.

Rizki, M. M., L. A. Tamburino, and M. A. Zmuda (1993) Evolving multi-resolution feature detectors. In *Proceedings of the Second Annual Conference on Evolutionary Learning*, eds. D.B. Fogel and W. Atmar, La Jolla, CA: Evolutionary Programming Society, 57-66.

Rizki, M. M., L. A. Tamburino, and M. A. Zmuda (1994) E-MORPH: A two-phased learning system for evolving morphological classification systems. In *Proceedings of the Third Annual Conference on Evolutionary Learning*, eds. A. V. Sebald and L. J. Fogel, River Edge, NJ: World Scientific, 60-67.

COMPUTER-AIDED FIXTURE CONFIGURATION DESIGN
USING ADAPTIVE MODELING LANGUAGE (AML)

Yiming (Kevin) Rong
Assistant Professor
Manufacturing Systems Program

Southern Illinois University at Carbondale
Carbondale, IL 62901-6603

Final Report for:
Summer Faculty Research Program
Wright-Patterson Laboratory

August 1995

COMPUTER-AIDED FIXTURE CONFIGURATION DESIGN
USING ADAPTIVE MODELING LANGUAGE (AML)

Yiming (Kevin) Rong
Assistant Professor
Manufacturing Systems Program
Southern Illinois University at Carbondale

Abstract

Flexible fixturing is an important aspect of flexible manufacturing systems (FMS) and computer-integrated manufacturing systems (CIMS). Modular fixtures are the most widely used flexible fixtures in industry for job and batch productions. Computer-aided fixture design (CAFD) has become a research focus in implementing FMS and CIMS. Fixture configuration design is an important issue in the domain of CAFD. A review of the current research in CAFD indicates that one major problem impeding the automated fixture configuration design (AFCD) is the negligence of study on fixture structures. This research investigates fundamental structures of dowel-pin based modular fixtures and fixturing characteristics of commonly used modular fixture elements. A modular fixture element assembly relationship graph (MFEARG) is designed to represent combination relationships between fixture elements. Based on MFEARG, an adaptive modeling language (AML) which is applying the objective-oriented programming technique is used to develop the AFCD system, including the core modules of fixture unit generation with alternatives and fixture unit placement into appropriate positions on a baseplate. A prototype system for AFCD with dowel-pin modular fixtures is presented in this report.

COMPUTER-AIDED FIXTURE CONFIGURATION DESIGN
USING ADAPTIVE MODELING LANGUAGE (AML)

Yiming (Kevin) Rong

Introduction

Reducing production cycle time and responding to rapid changes of product design is a means of surviving and thriving in the competitive market for most manufacturing companies. Manufacturing planning, including tooling, makes a major contribution in the production cycle. With the development of CNC technology which makes machining time much shorter than ever, attempt to reduce manufacturing time is focused on decreasing the time involved in workpiece setup. Flexible fixturing has become an important issue in flexible manufacturing systems (FMS) and computer-integrated manufacturing systems (CIMS) [1]. There are several different categories of flexible fixtures such as phase-change material, modular, adjustable, and programmable fixtures, where modular fixtures are the most widely used in industry [2]. Modular fixture configuration design is a complex and highly experience-dependent task. This impedes further applications of modular fixtures. Development of computer-aided fixture design (CAFD) systems is necessary to make manufacturing systems truly flexible.

Figure 1 shows an outline of fixture design activities in manufacturing systems, including three steps: setup planning, fixture planning, and fixture configuration design. The objective of setup planning is to determine the number of setups needed, the orientation of workpiece in each setup and the machining features in each setup. Fixture planning is to determine the locating and clamping points on workpiece surfaces. The task of fixture configuration design is to select fixture elements and place them into a final configuration to locate and clamp the workpiece. As more and more CNC machines and machining centers are employed, many operations can be carried out within a single setup, which needs to be ensured by a well designed fixture configuration. This research focuses on automated fixture configuration design (AFCD).

Fixture Design and Flexible Fixturing

Fixtures are important in both traditional manufacturing and modern FMS, which directly affect machining quality, productivity and cost of products. The time spent on designing and fabricating fixtures significantly contributes to the production cycle in improving current products and developing new products [3]. The primary requirement for a fixture is to locate and secure the workpiece in a given position and orientation on a worktable of machine tool in order to ensure the manufacturing quality. Locators and supporters are usually used in contact with locating surfaces of the workpiece to restrict six degrees of freedom, including linear and rotational motions. The locating surfaces may be plane, concentric internal, or external profile surfaces of the workpiece. Locating methods in fixture design include utilization of three planes (3-2-1 method), one plane and two holes, two planes and one hole, and long and short V-block [4]. To secure the workpiece on a fixture, clamps are utilized to keep a stable locating against the machining force, including vertical (top) and horizontal (side) clamping. To satisfy the primary fixturing requirement, locating accuracy and fixturing stability (equilibrium and rigidity) should be the main concerns in fixture design

Besides the primary requirement of fixture design, many other demands also need to be met, such as, ensuring productivity (e.g., easy load and unload of the workpiece, utilization of automated or semi-automated clamping devices, easy chip disposal), special design for reducing the deformation of weak-rigidity workpieces, simple and safe operation (e.g., the use of anti-mistake function components for costly workpieces), and effective cost reduction (considering fixture material and fabrication processes and using standard elements with priority). Because with a CNC machine tool, multiple operations can be completed in one setup, the fixture configuration design is restricted by a space availability for placing fixture elements and has to be well designed to avoid possible interference with NC path. Hence the fixture design is a complicated process. The application of these fundamental principles to an individual fixture design mainly depends upon the designer's experience in manual fixture design. Collection and representation of the knowledge from designer's experience is a crucial part in the development of computer-aided fixture design (CAFD) systems [5].

With the development of CAD/CAM technology, especially more and more CNC machine tools and machining centers are used in manufacturing industry, the trend of products is towards wide variety and small lot size. Since the product production cycle becomes shorter and shorter, manufacturing systems need to be flexible with rapid response to product design changes. As far as CNC machine tools are employed, usually only the NC program needs to be changed when the product design is changed. NC programming may take days even hours by means of a computer aided NC program system. The cutting tools have been highly standardized and can be purchased in market. Unless using flexible fixtures or existing fixtures, the overall FMS would not realize the real flexibility. Flexible fixturing is desired to adapt with the variation of product designs in FMS and CIMS. A number of different methods have been proposed for flexible fixturing, including flexible fixtures with phase-change material, programmable clamps, adjustable and modular fixtures [6].

Modular Fixtures and Computer-aided Fixture Design (CAFD)

Currently modular fixtures are the mostly used flexible fixtures in industry. Modular fixtures were originally developed for job or small batch production to reduce the fixturing cost, where the dedicated fixture was not economically feasible [7]. Modular fixture is assembled following combination principle by selecting the exiting standard elements. The flexibility is derived from the large number of fixture configurations from different combinations of the fixture elements which may be bolted to a baseplate [8]. Modular fixture elements can be disassembled after a batch of parts are produced and reused for new parts. The use of modular fixtures decreases the tooling cost and storage floor, especially shortens the lead time. It may take days even hours to build a modular fixture for a new product production. In comparison, it takes weeks even months to design and fabricate a dedicated fixture. The major difficulties of applying modular fixtures in industry are the complexity of fixture configuration design and verification which requires manufacturing knowledge and fixturing experience. The development of CAD/CAM techniques encourages the research on CAFD to automatically generate fixture configurations.

Currently three types of CAFD methodologies have been studied. One is to develop knowledge-based expert systems for the selection of locating methods, fixture elements, and fixture configurations [9]. The second approach is automatic fixture planning based on kinematic analysis and a series of design rules [10]. Basically, these researches are concentrated on fixture planning and cannot automatically generate fixture configurations. Since a good fixture design is highly dependent on designer's experience, the third approach utilizes the successful fixturing knowledge presented in exiting fixture designs to generate a new design. Group technology (GT) based CAFD systems have been developed for modular fixture design [11]. GT principle is applied to identify the similar fixture designs in a fixture design database. The most similar fixture design is provided to retrieve. Graphics functions in a CAD package are utilized to modify the fixture design for new parts. This is not an automated fixture design method, which makes use of the expert knowledge in existing fixture designs and specially valuable for complex fixture designs.

According to an analysis of fixture structures, a fixture can be decomposed into three levels, i.e., the functional units, fixture elements, and functional surfaces [12]. Once a fixture structure is analyzed, the fixture design can be described as a search for a match between the fixture structure and fixturing features of the workpiece.

Fixture Structure Analysis

Fixturing features of a workpiece have been analyzed, including geometric, operational, and fixturing surface information [4]. Fixture structure can be decomposed into fixture units, fixture elements, and functional surfaces, where the fixture units play a critical role in fixture configuration design.

A fixture structure is defined as a set or an assembly of fixture elements. Let F denote a fixture and e_i ($i = 1, 2, ..., n_e$) a fixture element in F, where n_e is the number of fixture elements in F, i.e.,

$$F = \{e_i \mid i \in n_e\} \tag{1}$$

This is a representation of a fixture at the level of fixture elements.

A fixture consists of several fixture units. In each fixture unit, all elements are connected one with another directly where only one element is connected directly with the baseplate and one or more elements are contacted directly with workpiece serving as locator, clamp or support. Let U_i denote a fixture unit in a fixture, we have:

$$U_i = \{e_{ij} \mid j \in n_{ei}\} \tag{2}$$

where n_{ei} is the number of elements in unit U_i.

Therefore a representation of a fixture at the fixture unit level can be written as:

$$F = \{U_i \mid i \in n_u\} \text{ and } F = \{\{e_{ij} \mid j \in n_{ei}\} \mid i \in n_u\} \tag{3}$$

where n_u is the number of units in fixture F.

A fixture element consists of several functional surfaces which can either serve as a locating or clamping surface in contact directly with workpiece or serve as supporting or supported surfaces in contact with other fixture elements. An element becomes:

$$e_i = \{s_{ik} \mid k \in n_{si}\} \tag{4}$$

where s_{ik} denotes the functional surface k on fixture element i; and

n_{si} is the number of functional surfaces the element i contains.

By combining formulas 3 and 4, a fixture can be represented at the functional surface level in the following form:

$$F = \{\{\{s_{ijk} \mid k \in n_{sij}\} \mid j \in n_{ei}\} \mid i \in n_u\} \tag{5}$$

In this research, four most popular fixture units are utilized in AFCD, i.e., vertical locating unit (VLU), horizontal locating unit (HLU), vertical claming unit (VCU), and horizontal clamping unit (HCU). To generate a fixture unit in AFCD, the acting height is the most important parameter and needs to be determined first. The acting height is the distance from the acting (contacting) point of the unit to the baseplate surface. Figure 2 shows the acting heights of different fixture units in a fixture design. In general cases, several fixture elements need to be assembled together to achieve the acting height.

<u>Fixture Element Modeling and Assembly Features</u>

Fixture configuration design is a process of selecting fixture elements from a fixture element database and allocating them together in space according to a certain sequence. In AFCD, a fixture element database needs to be built up, in which the

geometry information and assembly features are represented in a local coordinate system. The methodology of selecting fixture elements and assembling them together to form a fixture units is one key issue in AFCD. Studying on the assembly relationship between fixture elements and extracting basic combinations of the elements is a way to achieve automated fixture configuration design. In fact, the assembly relationships between modular fixture elements are not arbitrary but constrained. A fixture element can be only assembled with a fraction of other modular fixture elements and usually it can only be used in one or several units. In order to establish assembly relationships between fixture elements, assembly features of fixture elements need to be defined. Following functional surfaces are defined as assembly features of fixture elements: supporting and supported faces, locating, counterbore and screw holes, fixing slots, pins, and screw bolts. Figure 3 shows these assembly features. A supporting face is the surface that can be used to support other fixture elements or workpiece. A supported face is the surface that is supported by other fixture elements in a fixture design. A locating hole is the hole machined to a certain accuracy level and can be used as a locating datum with locating pins. Counterbore holes and fixing slots are used to fasten two elements with screw bolts.

In modular fixture systems, assembly features of elements are designed with standard dimensions. Other parameters of an assembly feature are the position and orientation of the feature in the element's local coordinate system, which are represented in a matrix form:

$$F = (V\ P)^T \tag{6}$$

where: $V = (\ v_x\ v_y\ v_z\ 0\ )$ is the homogeneous representation of the orientation vector V;

$P = (\ x\ y\ z\ 1)$ is the homogeneous coordinate of origin of feature F.

If F is a face type feature, its origin P is a point on the face, and the orientation vector V is normal to the face and points out from it. If F is a hole type feature, its origin P is the center of the hole end circle, and V points outwards along the axis of the hole. If F is a pin type feature, its origin P is the center on the tip of the shaft and V points outwards along the axis of the shaft. In the case of fixing slots, the origin P and vector V are defined as shown in Figure 3.

Fixture Element Assembly Relationship

In order to automatically select and generate fixturing units in fixture configuration design, the assembly relationships between fixture elements needs to be analyzed and represented in a computer compatible format. A modular fixture element assembly relationship graph (MFEARG) has been developed to represent the assembly relationships in building fixture units. Figure 4(a) is a partial MFEARG composed of real fixture elements, showing assembly relationships of the fixture elements for possibly building a VLU. An MFEARG can be defined, without loss of generality, as a directed graph (digraph) G, as shown as in Figure 4(b), i.e.,

$$G = (V, E)\qquad\qquad(7)$$

and $V = \{ v | v \in$ fixture elements$\}$; $E = \{ e \,|\, P (v_i, v_j) \wedge (v_i, v_j \in V)\}$; where V is a set of vertices representing fixture elements in a fixture unit; and

E is a set of edges representing the assembly relationship of fixture elements.

The edge $e (v_i \rightarrow v_j)$ presents that fixture element v_i, the start-vertex, can be mounted on the fixture element v_j, the end-vertex. The number of edges going to an end-vertex denotes an indegree of the vertex and the number of edges coming from a start-vertex denotes an outdegree of the vertex. An edge $e (v_i \rightarrow v_i)$ is called a self-loop if fixture element v_i can be assembled to its own kind. Utility Cube is one of such kinds.

A directed-path is a sequence of edges $v_{i1} \rightarrow v_{i2} \rightarrow v_{i3} \rightarrow \; ...$ such that the end-vertex of e_{i-1} is the start-vertex of e_i, which represents the possible assembly relationship for building a fixture unit. If the indegree of a vertex is zero (e.g., v_1, v_2, or v_3, in Figure 4), that means that no fixture element can be mounted on the fixture element. Locating tower is one of such kinds of fixture elements. Similarly the outdegree of v_8 is zero, which means there is no other fixture elements it can be mounted to except the baseplate. Therefore, a complete directed-path represents a possible formation of a fixture unit.

Fixture Unit Generation

In generating and selecting a fixture unit, all possible assembly relationships in building fixture units are presented in correspondence with MFEARGs. When the acting height of a fixture unit is input, a fixture unit generation module is activated to search all possible combinations and find out all fixture unit candidates which satisfy the acting

height. In the module, locators and clamps (the fixture elements directly in contact wit⁴ the workpiece) are first selected. Assuming a locator or clamp is selected as v_i, we get a sub-digraph G' of G:

$$G' = (V', E')$$ (8)

where $V' \subseteq V$ and $E' \subseteq E$

In G', v_i is the only fixture element with a zero indegree. All the directed path originally starts from v_i. Sub-digraph G' represents all possible fixture element assembly relationships as v_i is chosen as the locator or clamp. The process to generate a fixture unit becomes a search process in G' with an objective of finding the directed paths $v_i \rightarrow v_{j1} \rightarrow v_{j2} \rightarrow ... \rightarrow v_{jm}$ which satisfy the following acting height constraint:

$$H = h(v_i) + \sum_{k=1}^{m} h(v_{jk})$$ (9)

where: $h(v)$ is the acting height of fixture element v; and

H is the acting height desired for the fixture unit.

The fixture unit candidates may be listed in different sequences according to: 1) the number of fixture elements used in the fixture unit; 2) the total weight of the unit; and 3) the volume of the unit. When a specially high accuracy or stiffness is required, the fixture unit with the least number of elements is chosen with priority. In case a light fixture body is desired, the lightest fixture unit is first selected. If the spatial restriction becomes a big problem in the process of fixture unit mounting, the fixture unit with the smallest volume is the one to be selected.

Fixture Unit Placement

At the stage of fixture unit generation, only fixture elements are selected to satisfy the acting height. The exact positions and orientations of fixture units needs to be further determined at the stage of fixture unit placement. To place a fixture unit onto the baseplate, following factors are taken into consideration: the position and orientation of workpiece, the suggested locating or clamping points, the machining envelope, the position possessed by other mounted fixture units, and the positions of bushed and tapped holes on the baseplate. The placement requirements include a satisfaction of the acting point position of the unit to the desired fixturing point and the assembly relationship

between the fixture unit and the baseplate. Because the locating and tapped holes are distributed in a discrete manner on the baseplate, two parameters are used to indicate the positions of center of locating or tapped holes on the surface of baseplate, which are integers u and v in the ranges of (-N, N) and (-M, M). For the modular fixture system, the screws and holes are alternatively and evenly distributed in two dimensions (X and Y). The center positions of tapped holes on the baseplate can be represented parametrically as:

$$x_s = 2 T u + T ((v + 1) \bmod 2)$$

$$y_s = T v \tag{10}$$

The center positions of locating holes on the baseplate can be represented as:

$$x_h = 2 T u + T (v \bmod 2)$$

$$y_h = T v \tag{11}$$

where $u = -N, ..., -3, -2, -1, 0, 1, 2, 3, ..., N$;

$v = -M, ..., -3, -2, -1, 0, 1, 2, 3, ..., M$; and

T is a spacing increment between the tapped and locating holes in the row or column directions.

In placing a fixture unit onto the baseplate, a fixturing point (x^*, y^*, z^*) and direction is the target to be approached by the acting point and acting direction of the unit. The acting height of the unit is designed to approach the target in z direction, which is presented by Eq. 9. Therefore the fixturing point is projected onto XOY plane with the target (x^*, y^*). The two parameters are determined for the center position of the tapped hole on the baseplate which is nearest to point (x^*, y^*):

$$v^* = \mathrm{div}(y^* / T + 0.5)$$

$$u^* = \mathrm{div}((x^* - T ((\mathrm{div}(y^* / T + 0.5) + 1) \bmod 2)) / 2 T + 0.5) \tag{12}$$

The coordinates of the nearest tapped hole can be calculated with Eq. 10 where u^* and v^* are the variables. The determination of the center position of the locating hole follows a similar procedure and sometimes is not necessary when standard modular fixture elements are utilized because these holes are evenly distributed. The placement range of a fixture unit largely depends on the fixturing direction. Once the fixturing direction is specified, an acceptable placement range can be determined by considering the information of the fixture unit.

It should be noted that when the fixturing points are inconsistency with the desired values, the workpiece needs to be moved according to the actual locating positions. The positions of clamping units are usually adjustable in the clamping directions.

Object-oriented Programming and Adaptive Modeling Language (AML)

AML is a special designed object-oriented programming tool for concurrent engineering, which provides a paradigm to model and organize knowledge to integrate and automate the entire engineering cycle from the part conceptual design to its production. AML supports a modular underlying architecture. The AML object architecture and syntax is consistent throughout the system's object classes. With AML, the workpiece or fixture element models can be dynamically created or modified. When application condition changes, these models as well as their relationships (especially the assembly relationships in fixture configuration design) will adaptively changed where dimension-list structure is utilized within AML.

In this research, AML is applied to model the workpiece and fixture elements, the establish assembly relationships of fixture elements, extract locating and clamping information from the workpiece model, automatically generate fixture units, and automatically place the fixture units onto the baseplate. The prototype AFCD system developed in this research works in the following manner: 1) the locating and clamping surfaces and points are extracted from the workpiece model interactively from the screen; 2) based on the coordinates of locating and clamping points selected, the acting heights of fixture units are calculated by considering the least clearance between the workpiece and baseplate, which is usually required for a minimum height of machine tool operations; 3) the fixture unit generation and selection module is used to generate suitable fixture units according to the acting heights; 4) the position on baseplate is determined, which is suitable for a fixture unit placement; 5) the workpiece and clamping units positions are automatically finalized according to the actual positions of locating units, which is carried out by using reference-object functions; and 6) finally an interference checking module may be called to check whether the fixture unit at this position interferes with the machining envelope, the workpiece and other fixture units that have been mounted. If

interference checking is not passed, the fixture unit is adjusted to the next candidate position. In some cases, no candidate mounting position is acceptable. An alternative fixture unit will be generated to ensure that the final output be a collision-free fixture design. When the application condition changes (e.g., workpiece geometry or raising height changes), the fixture configuration would be automatically modified and re-generated if necessary. Figure 5 shows a sample output from the AFCD system.

System Integration and Future Work

The research presented in this report provides a foundation of further development of AFCD with the object-oriented technique, especially with an integration with the process design system currently developed in Wright Laboratory of Air Force. In the continuing research, the objective is to make the AFCD system truly automatic and adaptive. Following work is in plan:

1) Automatic determination of locating and clamping surfaces and points, which involves advanced geometric reasoning and accuracy analysis. Once this step is finished, the AFCD can be expected fully automatic and adaptive;

2) Establishment of a complete MFEARDB to include more fixture units, e.g. horizontal-vertical locating units and Vee-blocks (including half-Vee blocks), which refers to more complex unit structure and dependent relationships between units;

3) Dealing with more complex workpiece model, e.g., free form or curvature fixturing surface;

4) Interference checking and optimization on fixture unit generation and placement with artificial intelligence;

5) Selection of fixture type and generation of customer fixture elements;

6) Fixturing Cost estimation; and

7) Integration with the process design system.

Acknowledgment

Support of this research from Research & Development Laboratories, AFOSR Program Office, and Wright-Patterson AFB is very much appreciated. Special thanks go

References

1. C. H. Chang, "Computer-Assisted Fixture Planning for Machining Processes," Manufacturing Review, Vol. 5, No. 1, 1992, pp. 15-28.
2. Y. Rong, S. Li, Y. Bai and Y. Zhu, "Development of Flexible Fixturing Technique in Manufacturing," *Symposium on Manufacturing Science*, Evanston, IL, May 27-28, 1994, pp. 101-105.
3. G. J. Hess, "Best Practice for Manufacturing Excellence beyond CIMS," CASA/SME Conference, East Moline, IL, Dec. 10, 1992.
4. Y. Rong, J. Zhu and S. Li, "Fixturing Feature Analysis for Computer-Aided Fixture Design," *Manufacturing Science and Engineering*, ASME WAM, New Orleans, LA, Nov. 28-Dec. 2, 1993, PED-Vol. 64, pp. 267-271.
5. A. Y. C. Nee and A. Senthil Kumar, "A Framework for an Object /Rule-based Automated Fixture Design System, " *Annual of CIRP*, 1991, pp. 147-151.
6. B. S. Thompson and M. V. Gandhi, "Commentary on Flexible Fixturing," Applied Mechanics Review, Vol. 39, No. 9, 1986, pp. 1365-1369.
7. A. J. C. Trappey and C. R. Liu, "A Literature Survey of Fixture-Design Automation," Int. J. of Advanced Manufacturing Technology, Vol. 5, No. 3, 1990, pp. 240-255.
8. E. G. Hoffman, *Modular Fixturing*, Manufacturing Technology Press, Lake Geneva, Wisconsin, 1987.
9. D. T. Pham and A. de Sam Lazaro, "AUTOFIX - an Expert CAD System for Jig and Fixtures," Int. J. of Machine Tools & Manufacture, Vol. 30, No. 3, 1990, pp. 403-411.
10. Y. C. Chou, V. Chandru and M. M. Barash, "A Mathematical Approach to Automatic Configuration of Machining Fixtures: Analysis and Synthesis," J. of Engr. for Industry, Vol, 111, pp. 299-306.
11. Y. Zhu and Y. Rong, "A Computer-aided Fixture Design System for Modular Fixture Assembly," ASME WAM, *Quality Assurance through Integration of Manufacturing Processes and Systems*, PED-Vol. 56, Anaheim, CA, Nov. 8-13, 1992, pp. 165-174.
12. Y. Rong and Y. Bai, "Automated Generation of Fixture Configuration Design," ASME Transaction, Journal of Engineering for Industry (to appear).
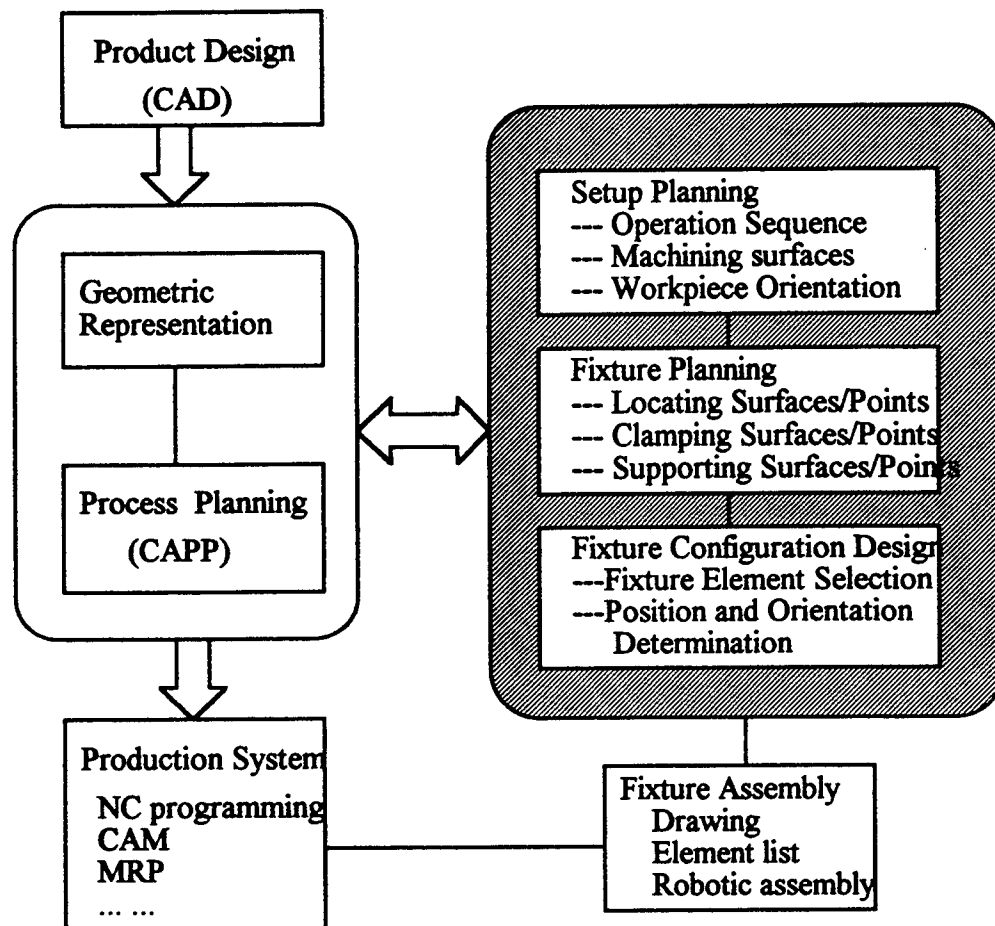
Figure 1. Fixture Design in Manufacturing Systems
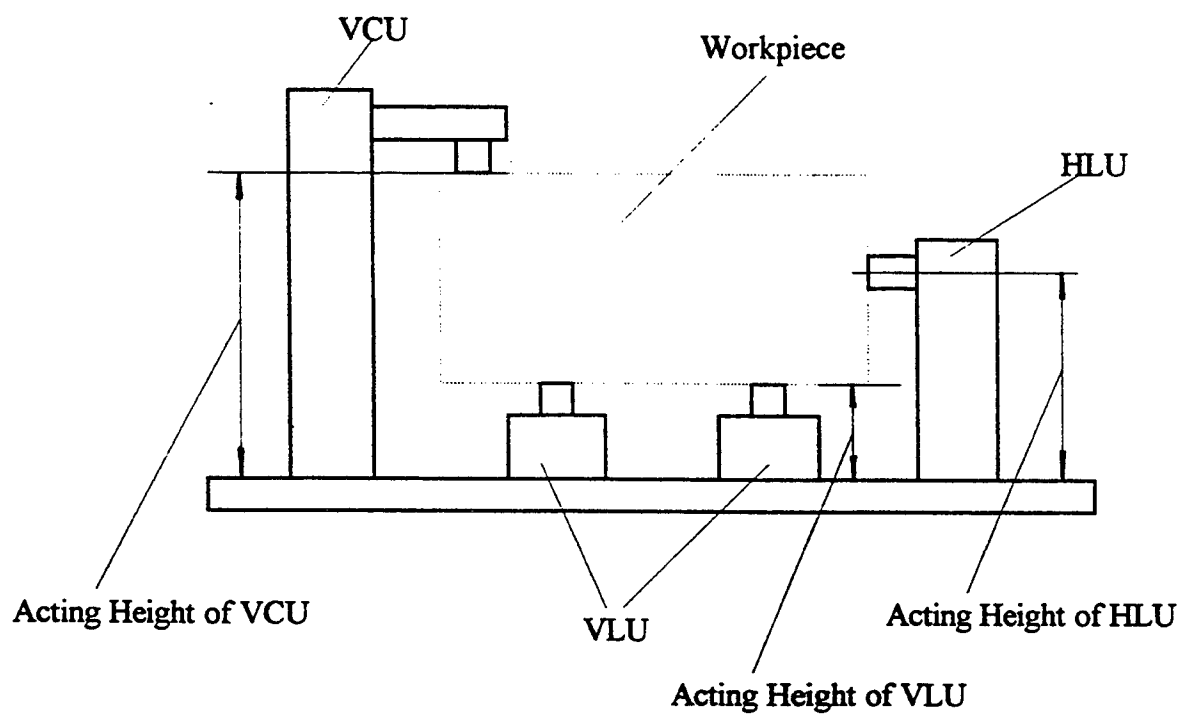
Figure 2. Acting Heights of Fixture Units

Supporting Face

Supported Face

Pin Hole

Counterbore Hole

Screw Hole

Pin

Screw Bolt

Fixing Slot

Figure 3. Assembly Features of Fixture Elements

Figure 4(a). Modular Fixture Assembly Relationships
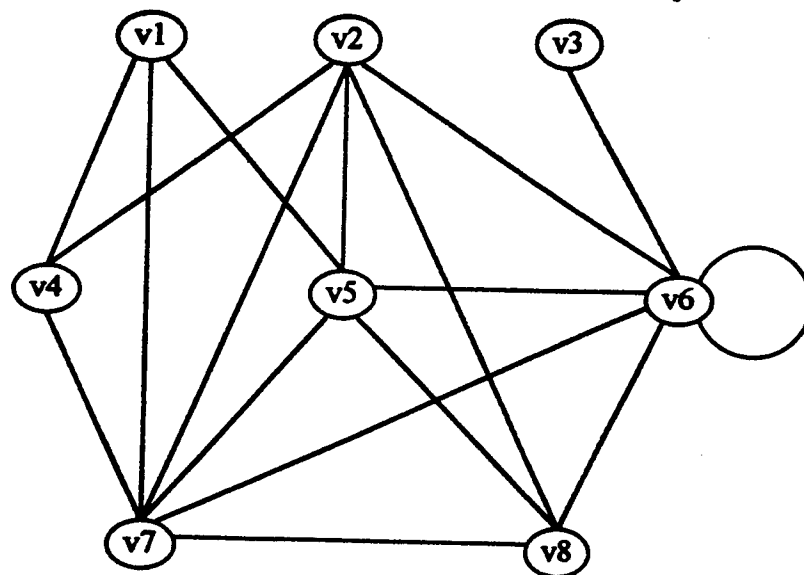
Figure 4(b). A Sketch of MFEARG Model
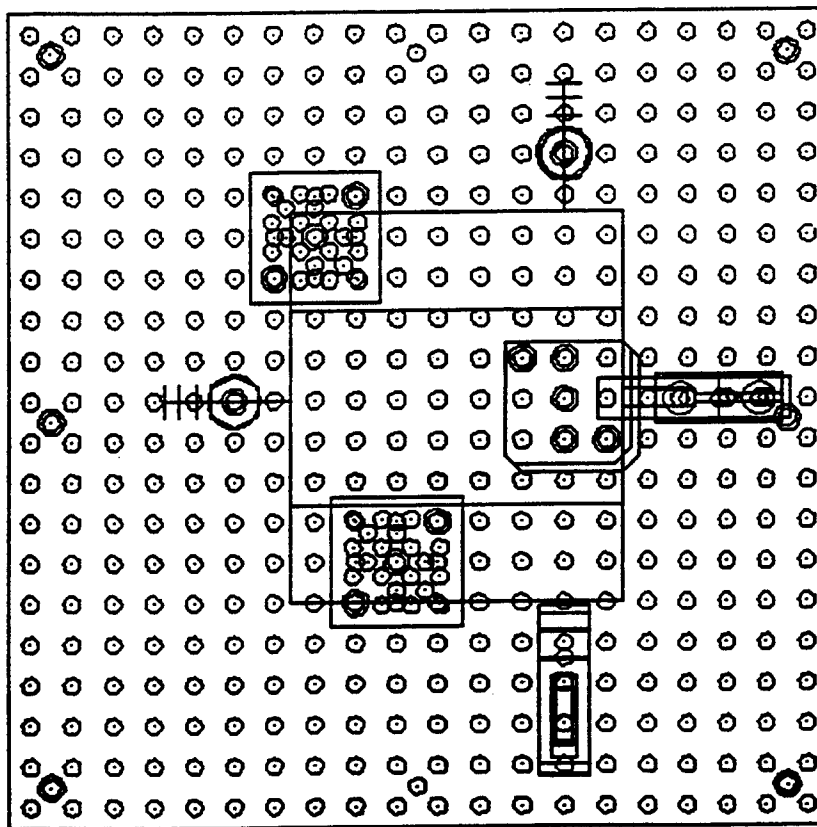
(top view)

Figure 5. Example of Fixture Design with AFCD
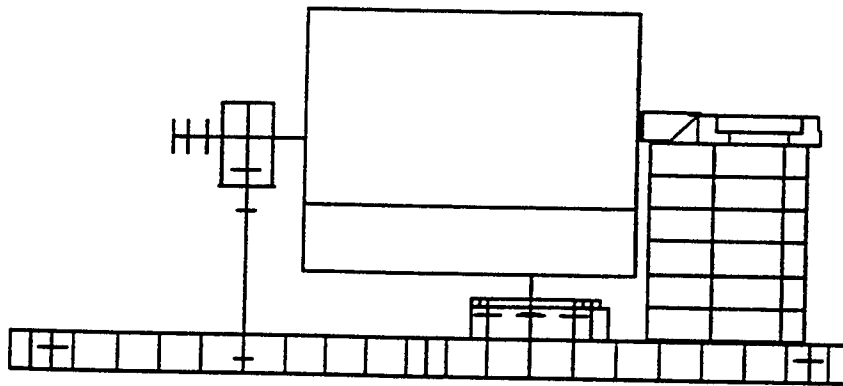
(side view)

ADA-SDP: AN ADVANCED AVIONICS SOFTWARE-DEVELOPMENT PROTOTYPE

Stuart H. Rubin
Assistant Professor
Department of Computer Science

Central Michigan University
Pearce Hall 403E
Mt. Pleasant, MI 48859

ADA-SDP: AN ADVANCED AVIONICS SOFTWARE-DEVELOPMENT PROTOTYPE

Stuart H. Rubin
Assistant Professor
Department of Computer Science
Central Michigan University

Abstract

The goal, of this on-going project, is to provide avionics systems designers with a tool (i.e., the Avionics Designer's Associate, or simply ADA), which assists software engineers in the rapid prototyping and testing of system models. A system model contains details of the planned system, but only to a level deemed adequate for integration testing.
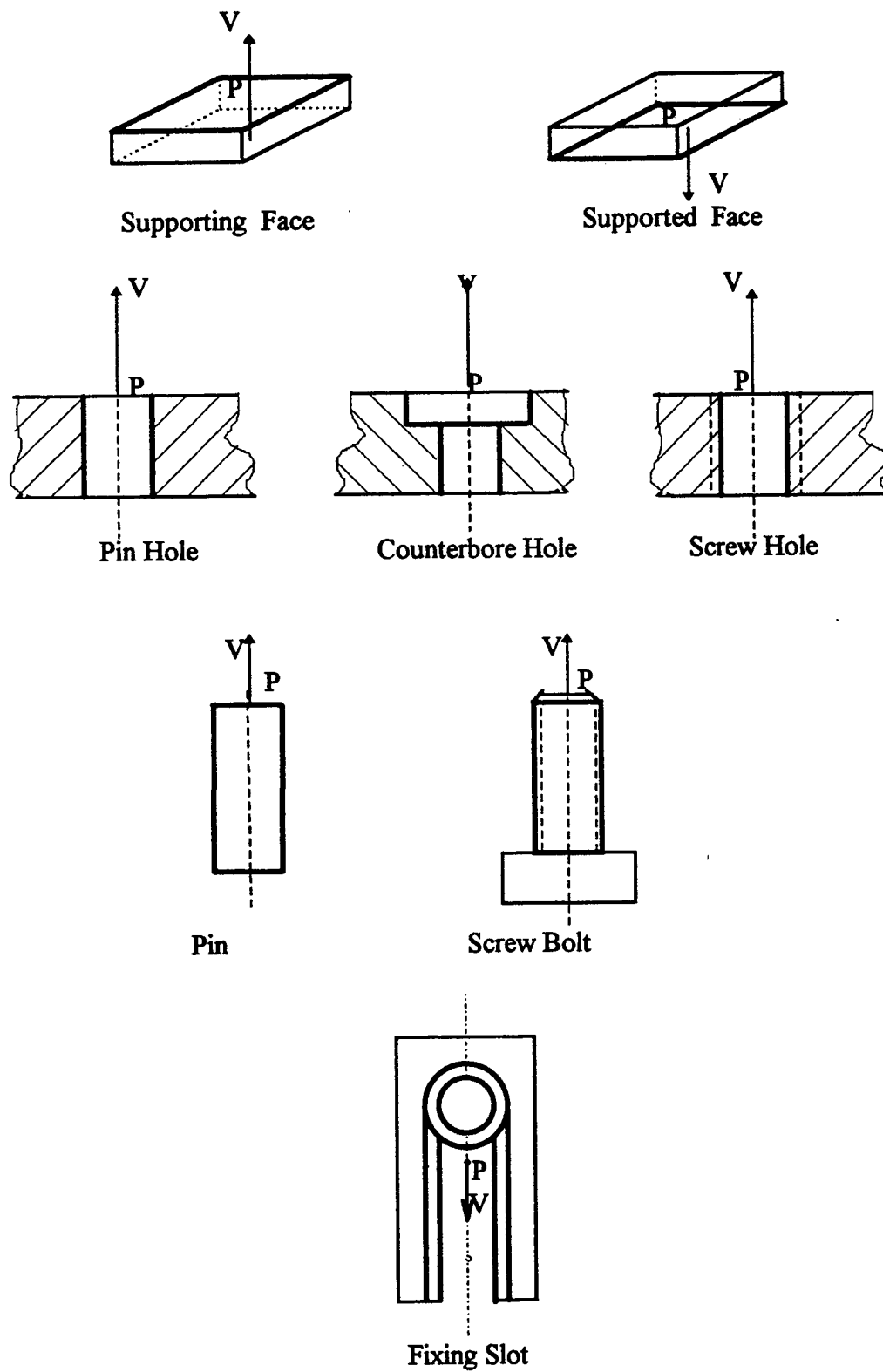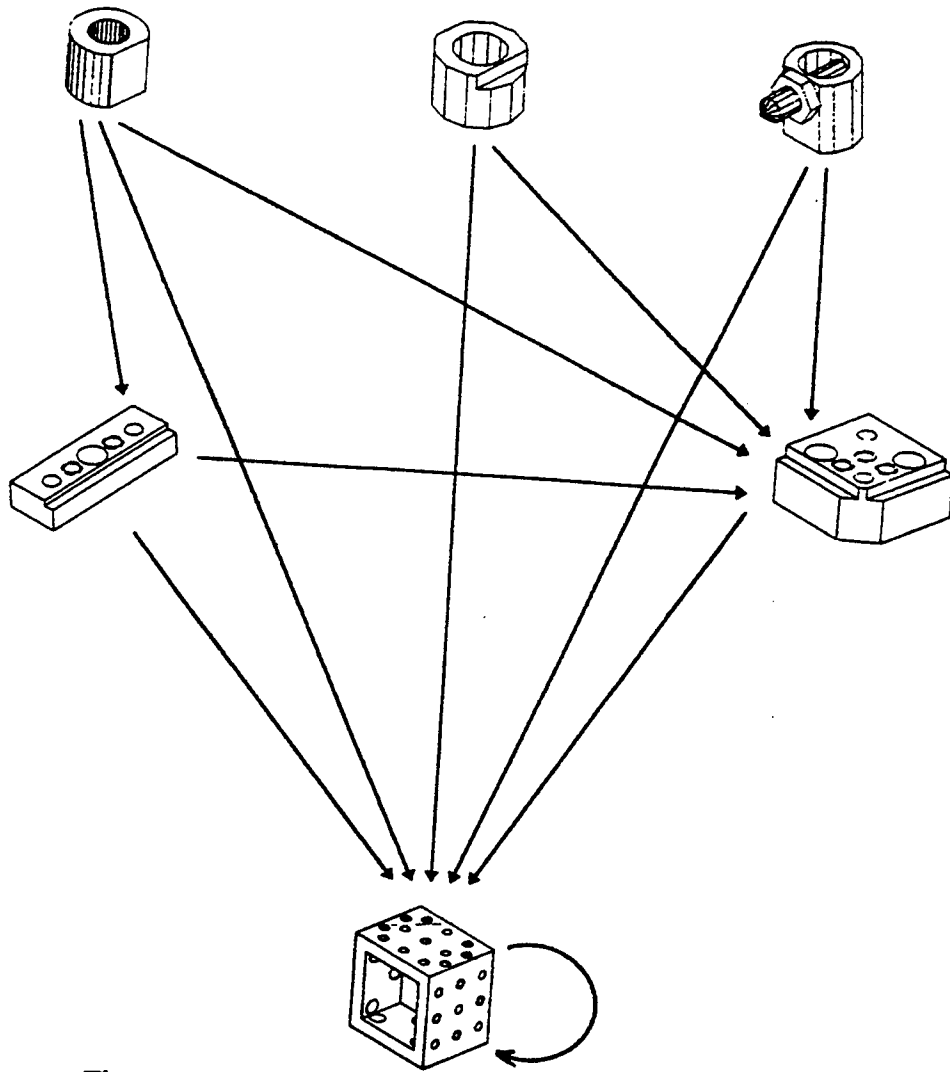
Models are executable prototypes. Modeling is closely tied to simulation, which refers to the exercise of a model over a variable parametric space. Model simulations not only provide the engineer with feedback pertaining to the validity of a proposed design, but additionally allow competing designs to be compared on one or more parameters (i.e., sensitivity analysis).

Models are defined from a base of several hundred primitive constructs. These constructs can define additional constructs hierarchically. All constructs (i.e., including their block names, icons, descriptive text, multimedia files, etc.) are placed in library folders in accordance with their operational domain.

This past summer, an $expert^n$--system was constructed, which retrieved software for reuse. This expert system is itself reusable and consists of many sub-systems -- any one of which can invoke any other. A key feature is that any $expert^n$--system need never be modified, for purposes of reuse, once saved in a repository. Rather, it communicates all information back to the caller and lets the caller decide how and when to use it. Thus, blocks in an $expert^n$--system have very low coupling (i.e., no off-model connections). In addition, $expert^n$--systems are, as their name suggests, organized in a hierarchy. This means that very complex decision-making systems can be called into play with minimal effort. Growing the repository is equivalent to learning. A two-hour video tape was delivered to Wright Laboratory, which depicts the use of the $expert^n$--system in retrieving software based on ambiguous specifications.

Table of Contents

ADA-SDP: AN ADVANCED AVIONICS SOFTWARE-DEVELOPMENT PROTOTYPE

Stuart H. Rubin

Introduction

The cost of digital avionics hardware has been decreasing exponentially since the introduction of the integrated circuit more than three decades ago. Unfortunately, the cost of supporting software has at best decreased linearly (and in some cases has actually increased). The reasons for this discrepancy are now clear. It seems that hardware is amenable to block reproduction (e.g., using e-beam machines to fabricate gate arrays), while software must be more or less individually tailored for each application. Reuse, where it occurs, is often minimal. In fact, even if 90 percent of the code can be reused, a majority of the overall development time will be spent on just 10 percent of the code. This seeming paradox is well-known. It is attributed to the difficulties associated with software maintenance and debugging. The problem is that these tasks are context-sensitive. Furthermore, understanding the context in which code occurs is not amenable to automation. It necessarily involves human factors {4, 10, 11, 12, 13, 20}.

This past summer, a study was undertaken to address the so-called software bottleneck. More specifically, it was desired to investigate the application of artificial intelligence technologies to breaking the software bottleneck. It was found that this may be accomplished through at least two routes: (1) simulating large-scale software systems for rapid prototyping and development {11}, (2) cataloging and retrieving hierarchical software blocks for reuse. Additional methods, while possible in theory, have so far proved impractical. They include automatic programming for general domains, logic programming in general domains, and 5th-generation languages over general domains. These methods are successful where the operational domain is much restricted a priori. This coincides with the maintenance of a reusable component library. The reason is simple.

Libraries are built from used components. By definition, components do not get built for a general domain -- only for a specific one. Thus, it can be shown that the class of automatic programming methods, in the large, coincides with the class of reusable-component methods in the large. One further stipulation deserves mention. That is, the argument only holds where the programming languages are context-sensitive or higher (i.e., 4th generation languages or higher). Thus, 3d generation languages, such as ADA, can only be improved so far (and there is still room for improvement). However, a theoretical point is reached where reuse methodologies (or equivalent) then dictate the level of language possible. This is a fundamental realization that is just beginning to be realized.

Reusable databases of software components and reusable knowledge bases of rules for information processing go together. This goes beyond the apparent structural similarity. There is also a functional necessity here.

Software retrieval is dependent on a highly cohesive characterization. The characterization of software, in turn, is dependent on its representation, which implies the use of domain-specific languages {3, 5, 6, 10, 15}. Such characterizations are amenable to domain-specific retrieval using an expert system technology. Indeed, a finer point is that since these technologies are dependent on domain-specific representations too; an expert system can be constructed for the retrieval and reuse of the specific knowledge necessary for its development. This was demonstrated last summer at Wright laboratory as detailed below.

Figure 1 depicts a screen capture of a hierarchical model having three levels of hierarchy, or H-blocks. Each level represents a separately compiled block module, which is stored in an appropriate repository (e.g., the ADA library). These blocks may be inserted into the model at any point using a simple click-and-drop operation. It is just as simple to delete them. Each block has its own local Start Model and Update Model buttons (since they were separately compiled). Only the buttons on the zeroth level will be active however. Start Model is used to begin operation, while the Update Model button is used to adjust the model for block acquisitions and deletions.
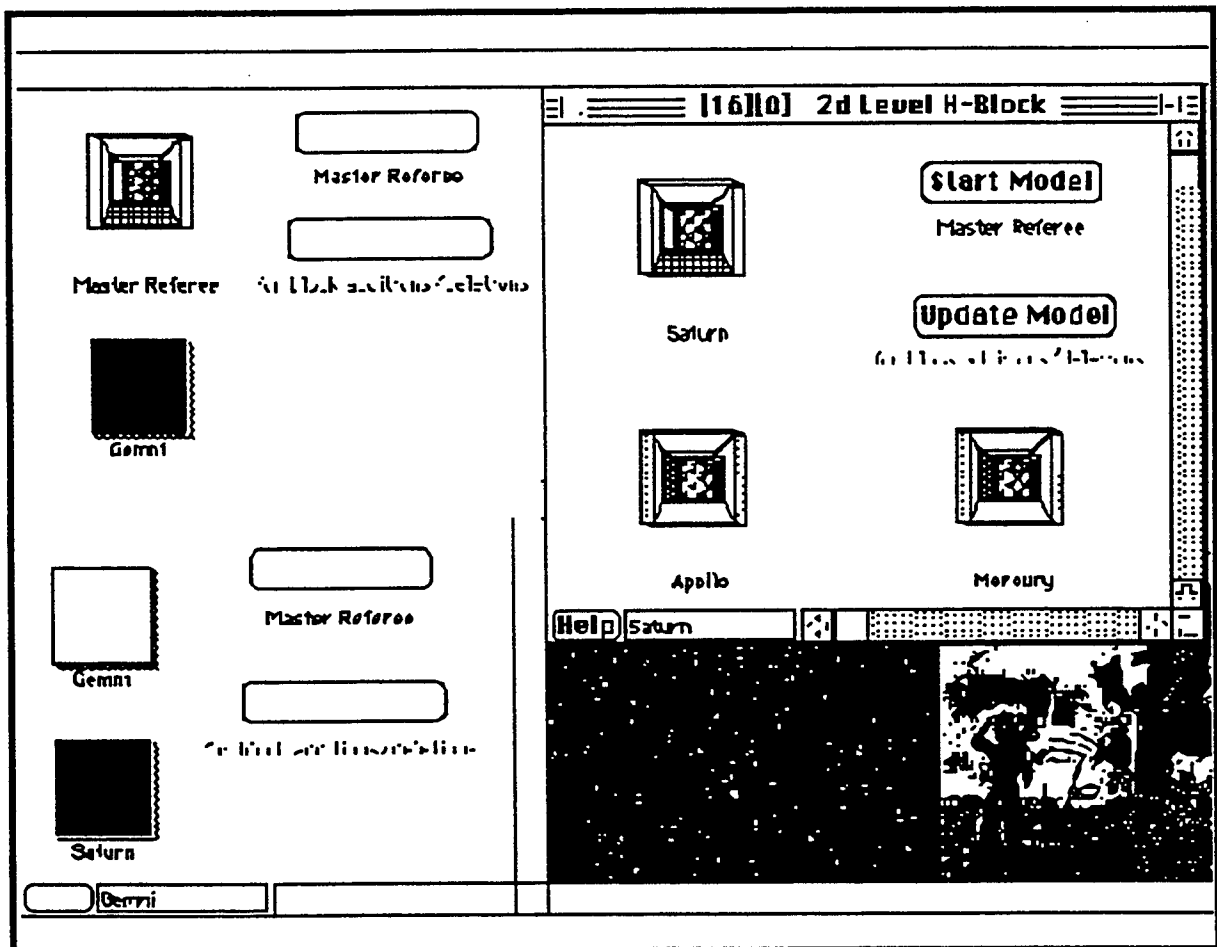

Figure 1. Hierarchical Model Structure

Figure 1 depicts three levels of hierarchy. The label given each H-block must correspond to that of the master referee on each level. An H-block can only be opened by calling this block label. A block can only be closed by using the call-back command, since outer levels are invisible within an H-block.

Methodology

One of the problematic areas that arises in the construction of any software repository for reuse is that of its effective use. It is not sufficient that the repository work in theory. The repository must be capable of retrieving code in accordance with the user's specifications. This is quite a different problem from that encountered in database design and implementation. Databases use a query language such a SQL, DBIV, or equivalent. The user specifies a retrieval based on matching a set of attributes and relations. For example, it is quite a different problem to retrieve all part numbers matching the template say, "95-*1*-3" than it is to retrieve all software modules that will invert a positive-definite matrix. Notice the generality of the *syntactic* characterization language, in the former case, and the necessary specificity of the *semantic* characterization language in the latter case.

Syntactic retrieval can be effected by an algorithm. Semantic retrieval can only be effected by a rule base. Rule bases are necessary because they are pliable and readily amenable to modification; whereas, algorithms are rigid and only assured to operate correctly once debugged. In other words, the problem of software retrieval requires specific characterization and machine learning in its solution. This assertion too has a translation as follows. Software reuse, in the large, requires an extensible characterization; or, in other words, one must write software to retrieve software (i.e., there are no fixed points)! This general case goes beyond the experimental results achieved this past summer. Nevertheless, our experience in coding leads us to agree with the general result. For example, while belief sliders (see below) are used to represent fuzzy certainties, there are times when we wish that they would represent engine temperatures, etc. instead. Fixing a representation gains power at the expense of generality; and, our results this summer serve to verify the validity of this statement.

The reader should, by this point, be left wondering how to pursue the goal of this project. After all, writing software to retrieve software to write software to retrieve software... sounds like a little bigger sandwich than we would care to bite. Experimental results evidence no fewer than two solutions. First, and most evidently, the domain may be restricted to be theoretically trivial, although of practical significance. Here, characterization languages may be fixed and retrieval methodologies may even be rendered algorithmic. In effect, a knowledge-based retrieval of software for reuse is reduced to a database retrieval. Most software reuse systems today (the ADA language included) employ a variant of this basic methodology.

Second, a model of the software system may be rapidly prototyped for testing and evolutionary improvement. Here, the model is hierarchically constructed using software stubs (which, unlike the full-blown code, are readily modified). The final stubs then serve as specification descriptions for retrieval of the desired software from the repository. The stubs mix executable and non-executable descriptions. Retrieval is rule-based. This was the approach taken by our faculty summer research project.

☒ Begin Software Retrieval
Comment 0: Make sure that
the above checkbox is
checked and set the belief
slider to 1.0.

☐ Location of Apollo Known
Comment 1: This refers to the
library location of the
reusable Apollo software as
depicted by the video.

(a) The Master Referee Subsequent to User-Initialization

☐ How It Works
Comment 0: Listen for a
comment on how the
software is effectively
retrieved.

☒ Execute the Apollo Software
Comment 1: This software
will only be executed if the
belief slider is set to a value
greater than 50 percent.

(b) The Apollo Block after being Initialized by the Master Referee

☐ Begin Software Retrieval
Comment 0: Make sure that
the above checkbox is
checked and set the belief
slider to 1.0.

☒ Location of Apollo Known
Comment 1: This refers to the
library location of the
reusable Apollo software as
depicted by the video.

(c) The Master Referee after Call-Back from the Apollo Block

☐ How It Works
Comment 0: Listen for a
comment on how the
software is effectively
retrieved.

☒ Execute the Apollo Software
Comment 1: This software
will only be executed if the
belief slider is set to a value
greater than 50 percent.

(d) The Apollo Block after the Second Call by the Master Referee

☐ Begin Software Retrieval
Comment 0: Make sure that
the above checkbox is
checked and set the belief
slider to 1.0.

☐ Location of Apollo Known
Comment 1: This refers to the
library location of the
reusable Apollo software as
depicted by the video.

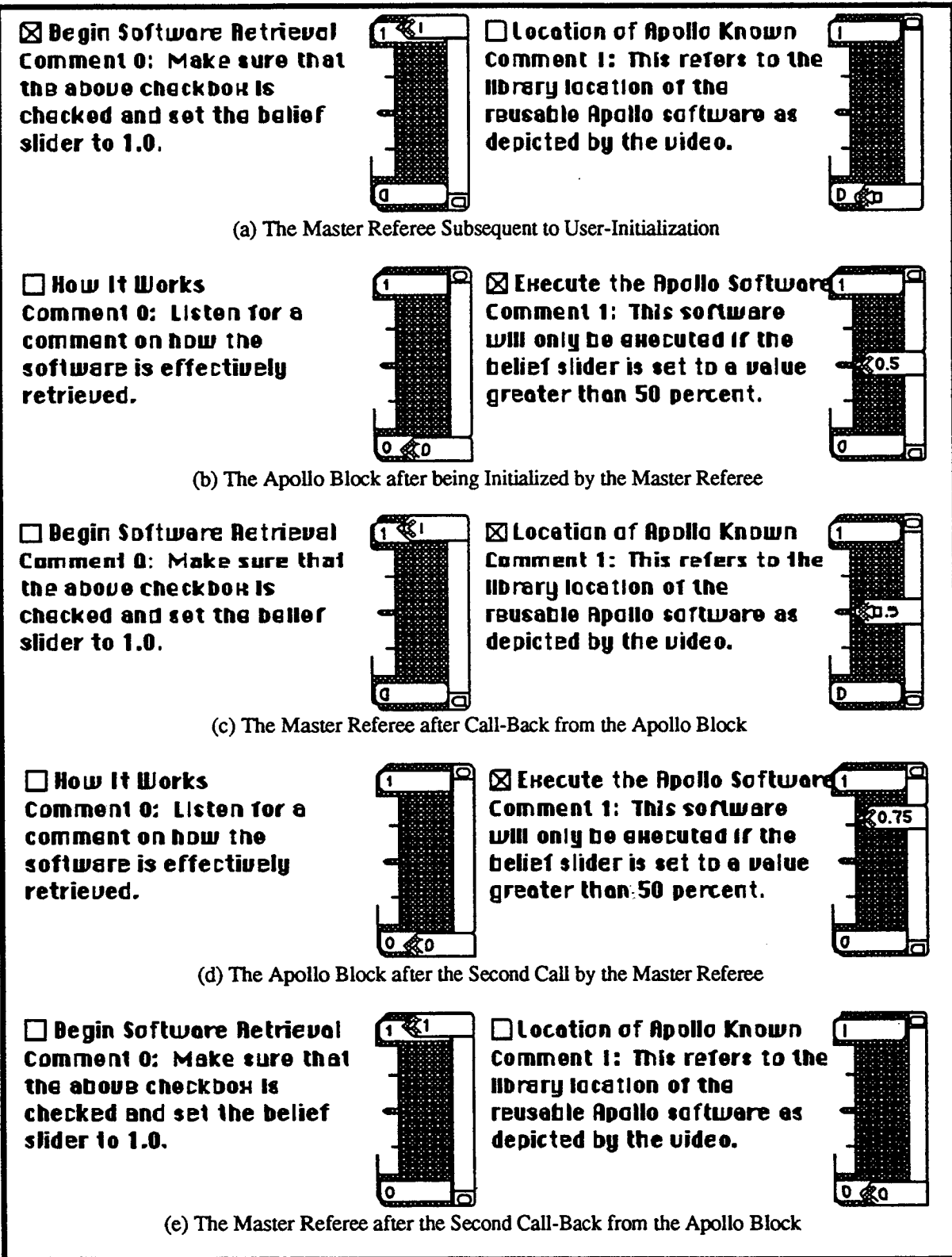(e) The Master Referee after the Second Call-Back from the Apollo Block

Figure 2 Knowledge Amplification in an Expert^{n}-System

Various simulation packages already exist for the purpose of functional modeling (e.g., GASP, GPSS, SLAM *et al.*). However, all existing packages exclusively rely on the creativity of the designer for their qualitative solution {2, 6, 8, 18, 19}. The problem here is that the design decisions of an expert are not captured for future application {1, 7, 12}. This work promises to improve design quality while reducing design costs {22, 23, 24}.

The summer project emphasized the construction of an expert^n-system, which the user can instantiate with rules for finding and retrieving software blocks from repository libraries {15, 16, 17}. The rules begin by having the user check the stub characterizations and the relative degrees of certainty associated with his or her selections. Up to 24 characterizations and beliefs per block may be entered. One purpose, served by the rules, is the translation of stub descriptions into corresponding software block structure names for retrieval. The block structure names may be found in the specified libraries. This then completes the cycle from rapid prototyping and testing to finished software product.

Figure 2 shows how relevant stub descriptions and their attendant beliefs have been set by the "Master Referee" and "Apollo" blocks in an alternating feedback loop. Such feedback loops, when not deleterious (see below), are useful for amplifying user-supplied knowledge for use in software retrieval or modification. Note that the first belief slider in the master referee is left at 100 percent certainty to make it slightly easier for the user to re-run the model. Notice that no further rules will fire automatically on conclusion.

One problem area that was not previously addressed will be addressed now. That is, what happens in the event that the desired software block is not found, but the rules locate a similar block using some rule-based metric? Here, the user will need to more or less manually modify the retrieved block. This can be quite arduous depending on the nature of the retrieved block. Reducing the granularity of the stored blocks tends to help; but, this is not always possible or practical. Nevertheless, reducing the granularity of the software is sufficient in the limit. Also, in keeping with the definition of randomness {15, 16, 17}, it is not possible -- even in theory -- to automatically modify every retrieved block in a directed way. After all, the reader will recall that this is the very reason for the reuse of high-level software in the first place! In the event of a non-exact match, ADA-SDP reports a non-matching condition and leaves the next step to the user.

Now, as mentioned above, the designed expert system is an expert^n-system. This means that specifications and beliefs that are unknown to the user, or too difficult to answer, may be "out-sourced" to a companion expert system. The companion expert system may be part of a very complex hierarchical arrangement, or it may be a single block. Companion expert systems may also out-source and so on. Results are recursively returned and processed by each expert system. Figure 3 depicts part of the ModL script for rule number zero in the master referee, which calls the Apollo block. Most of the depicted functions are not part of ModL itself, but were written as extensions for the inference engine. In particular, notice how specifications and beliefs are set in the next block and upon return to the master referee. The full rule-specification language (RSL) description is given in the appendix.

There are many details pertaining to the operation of the expert^n-system. Only the more salient ones will be presented here. For example, expert systems are arranged in hierarchical blocks. A block may call any other

```
** RULE 0:
        if      (NowInitializing)
                IF (spec0)          ** AND specj AND speck ...
        {
                CF = belief0 [2];   ** * beliefj [2] * beliefk [2] ...
                ThisisRule (0);

        }
                else; else if (NowFiring == 0)
        {
                if (NOT CallBack)
                {
                        ShowQuickTime ("snd001.mov", 1.0, FALSE);   // movie classical music
                        ShowQuickTime ("Kennedy.mov", 1.0, TRUE);   // why go to the moon
                        Speak ("We next will call the Apollo block, which will find the location of the "+
                            "Apollo software.");
                        UserError ("Rule 0:  We next will call the Apollo block, which will find the location "+
                            "of the Apollo software.  The level of certainty is "+certainty [FiredRules]+" percent.");

                        if (OutToFile AND NOT Canceled)
                        {
                                FileWrite (filenum, "Rule 0:  Looking for Apollo in ADA Library.", "", TRUE);
                                FileWrite (filenum, "         The level of certainty was "+
                                    certainty [FiredRules]+" percent.", "", TRUE);
                        }
                }

                CALLABEL ("Apollo");

                SET_NEXT_SP_BE (1, TRUE, 0.50);   // spec1 set to true; belief1 [2] is 0.50 in Apollo

                SET_RETURN_SP (0, -1);            // Set this spec0 to the not of returned specification [1].
                SET_RETURN_SP (1, 1);            // Set this spec1 to that of returned specification [1].
                SET_RETURN_BE_R (1, 1, NA);      // Set this belief1 [2] to the returned probability [1].

                Firenext ();
                Return (FALSE);                  // do not call-back
        }
```

Figure 3 Sample Rule Script

block label on its level; but, it may not call another block label on an outer level or an inner level. Outer levels can only be reached by return from the called block. Inner levels are reached by calling the hierarchical block label, which must correspond to the label of the master referee for exactly one block on the next level. Block labels may not be repeated on the same level. These enforcements will be recognized as analogs to structured programming. They serve to prevent the user from writing difficult to follow calling sequences (e.g., spaghetti code).

The ADA-SDP system not only enforces a structured development methodology; but, it also facilitates the debugging of any rule base while it is running. Compiled block script replaces previous faulty script. A special browse and return feature allows the user to investigate dialog boxes (Figure 4) -- including block specifications and beliefs and then allow the system to automatically return to the active block, opening and closing the

necessary blocks along the way. The Manual Back button is useful for debugging the script for processing call-backs. The table of local block labels alphabetically presents all blocks on the same hierarchical level as the current block. This information facilitates the development of the CALLABEL and associated statements.

Figure 4 The Default Dialog Box

Blocks are animated to show, using color-coded text, the calling block, if any (top), and the called block, if any (bottom). This information is useful for debugging purposes. For example, Figure 5 shows the animation

for two blocks that are locked in a cycle of calling each other. Notice how easy it is to ascertain this fact using the block animation. Again, repairs are facilitated by the compile-and-go environment. Note that not all cycles are deleterious (e.g., Figure 2).

Apollo   Master Referee   **Start Model**

                 Master Referee

Apollo   Master Referee   **Update Model**

 Master Referee   Apollo   for block additions/deletions

The Master Referee is used to launch the other blocks.
It serves as the external controller for the model.

Figure 5 The Use of Block Animation in Cycle Detection

The ADA-SDP system provides for two types of error handling. First are the catastrophic system errors (e.g., duplicate block labels, constant 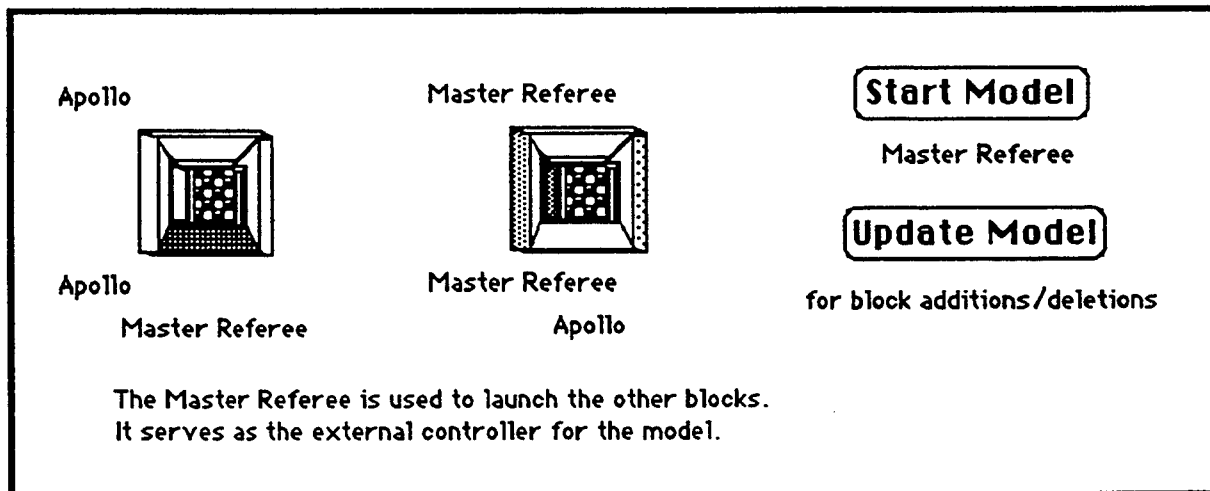value too small, etc.). These errors are flagged and require the user to modify the block structure or various parameters in one or more include files. Second are the rule-based errors (e.g., you cannot use return (true) inside of the master referee block). These errors instruct the user on what to repair and provide for a soft failure. Here, the user need only click on the start model button to begin anew subsequent to rule-base repair.

<u>Results</u>

Results in the ADA-SDP system are qualitative rather than quantitative. It would make little sense to plot the percent of software modules retrieved correctly when for example, it is known that this percentage is almost totally dependent on the operating domain. Thus, work this past summer has resulted in an operational expert^n-system, but preliminary results do not permit meaningful quantitative comparisons at this time, save one as follows.

A fuzzy logic-based certainty metric is assigned to each belief in the system. This reduces the number of specifications and belief sliders necessary to retrieve a given software block. For example, consider two specifications and their associated sliders: namely, the engine is running hot and the engine is running cold. Fuzzy logic allows these two sliders to combine their evidence to represent the whole spectrum of temperature qualifications. For example, the concept that an engine is running warm could be represented by a 60 percent

belief that it is running hot combined with a 40 percent belief that it is running cold. Similarly, the concept that an engine is running cool could be represented by a 40 percent belief that it is running hot combined with a 60 percent belief that it is running cold, and so on.

The expert^n-system works as predicted. The knowledge engineer programs rules and the appropriate clauses. The user selects the known specifications and the associated beliefs. The rules proceed to amplify this knowledge basis and, if under-determined, query the user to elicit the correct and needed knowledge. If the needed knowledge cannot be elicited, then the system admits that it cannot help to locate the requisite software module.

Four unique conflict resolution strategies serve to give the user control over the operation of the system. The user can choose on a block-by-block basis, or for every block, to resolve conflicts in decreasing order of a fuzzy-certainty metric, in increasing order of this metric, in random order (i.e., where the user chooses the number of standard deviations of randomness), or in rule-based order (i.e., for those who love to program procedurally). Naturally, the default order is in decreasing order of certainty. These orders become especially significant when it is realized that one expert block can call another and the order of calls is strictly dependent on the conflict resolution strategy, or combination of strategies, used.

Conflict resolution in decreasing order of certainty implies that the most probable rule will be first to be fired. Conflict resolution in increasing order of certainty implies that the least probable rule will be first to be fired. This ordering is rarely used except to answer the question: "Which software block is the opposite of what I am looking for?" For example, if you are searching for a block, which sorts numbers in decreasing order and nothing is found, then you might try looking for a block, which sorts numbers in increasing order by re-running the system using the increasing order of certainty option. Conflict resolution in random order (i.e., within 3 standard deviations of true random) is useful for browsing the software repository for blocks than may be similar enough to the sought-after software to justify manual modification. Rule-based order is quite a useful option where the order of the script dictates the order of firing. This order is useful for experienced algorithmic programmers.

The first three orderings are achieved using a recursive implementation of the quicksort algorithm. While this algorithm has an average complexity of $O(n \log n)$ in the average case, it has a worst-case complexity of $O(n^2)$ in the event that the vector is already sorted. Such a state of affairs can occur, for example, if the user were to set all belief sliders to 100 percent. Note that if a belief slider is set by the user or another rule to zero, then the corresponding specification will be automatically unchecked for the obvious reason. The fuzzy-certainty vector is always randomly permuted to within one standard deviation of true randomness and the resulting vector is passed on to the quicksort algorithm. This insures both an average and worst-case complexity of approximately $O(2n + n \log n)$, where the constant is affected by the complexity of the pseudo-random number generator (e.g., linear congruential), which is called. The uniform response produced is more appropriate for our purposes -- even at the cost of some additional CPU time, on the average.

The main results, of our summer research, pertain to new directions for reducing the cost of developing software. When you stop and think about this, it becomes obvious that the solution here cannot be overly simple, complex, or mathematical -- simply because all of these approaches have failed in practice. Again, the purpose of our research is to find a software development methodology that yields more reliable code in less time than would otherwise be the case. The results, which stem in part from the supplied video, follow. They are not meant to limit further innovation, but rather provide a framework within which subsequent innovation may operate.

- Choose as high a level language as the operational domain permits. Reuse will not be practical for languages on a lower level than the 3d generation languages (e.g., the ADA language), since lower-level languages do not realize the full potential of a domain-general context-free grammar. Type 0 (i.e., including 2-level grammars) and type 1 grammars, while theoretically desirable, are practically impossible to program in. They may be replaced by 4th and 5th generation languages if appropriate. Note that the stated methodology subsumes a 5th generation language. This does not give rise to contradiction.

- If the programming problem is trivial, then you are done. Otherwise, an executable prototype is built using an appropriate simulation language (e.g., ModL and its derivatives) {10, 21}. One of the features of this simulation language is that it permits executable and non-executable software stubs to be built. The purpose of the non-executable text is to serve as a specification, which is linked to the full-blown code in a specific repository. It is necessary, to maximize reuse, that the code, like the stubs in this respect, be as fine-grained as possible. The only exception has to do with software optimization, where a sequence of fine-grained modules (descriptions) is replaced by a larger, more-efficient module.

- Run the software simulation until you are satisfied that the model operates in accordance with specifications. Testing here refers to block integration -- not the testing of individual blocks. That is, the blocks represented by the stubs are assumed to operate correctly by virtue of being saved in the software repository. The same holds true for hierarchical blocks.

- Run the expert^n-system over the block stub descriptions (executable and non-executable -- depending on design) to find the library and block names of the corresponding full-blown code. Additional information may also be provided (e.g., limits, number of times tested, etc.). If an incorrect block is retrieved, or in some cases if no blocks are retrieved, the knowledge engineer must edit the expert^n-system. This system, like the software it is meant to retrieve, is modular in the sense that it is built from blocks and hierarchical blocks. If a block or hierarchical block is placed in a

library, it is assumed to be debugged and complete. It is augmented by being called by other blocks. Thus, the script need not be altered unless a bug surfaces. In fact, augmented blocks recursively define hierarchical blocks. There are two reasons that no blocks may be retrieved. First, if the block exists in the repository, but is not found, the rule base(s) need to be updated (i.e., truth maintenance). Second, if the block (or a similar block using different conflict-resolution strategies) is not found, then this block may be *random* relative to the stored repositories. Here, the programmer needs to write, test, and save the code -- but only once in view of reuse. Also, the fine granularity of the code will facilitate scripting all but the optimization blocks (i.e., after all a perfect system must have its loopholes).

• Hierarchically blocking an expert^n-system was described above. The expert^n-system is thus seen to be constructed out of reusable components. The summer video tape demonstrated that such a system is indeed capable of retrieving its own components for its construction. This is significant because it means that all models can then be bootstrapped, or knowledge acquisition is accelerated, if you prefer. Self-reference (i.e., in ADA-SDP) is the theoretically most demanding test possible {9, 14}. Our preliminary AFOSR proposal promised to prove the feasibility of this self-referential concept. We now consider it to have been proven possible. Now, just as knowledge reuse has been shown to be eminently practical, model (i.e., application software) reuse is too. We hope to demonstrate this concept in a few months. It will be detailed in a forthcoming SREP proposal. The concept is actually quite simple as follows. If one works in a narrow domain of expertise and constructs many software models in that domain, then there will be much overlap and duplicated work. Here, the idea is to build models from reusable components, which are retrieved from a software repository similar to that used for the knowledge bases. Furthermore, blocks and hierarchical blocks, in the repository, may be paired with full-blown coded implementations for immediate software realization. Again, all such blocks are assuredly high quality. When the expert^n-system and its capability for automated retrieval is factored in we truly have a learning automated high-level programming system. These results, while recursively enumerable by virtue of our summer program, are not recursive since we cannot prove that there is not a better way of doing things (e.g., simulated evolution). Thus, akin to debugging complex code, our proposed methodology must stand the test of time and evolve with it. Another note should be added as follows. The first compilers were rule-based. They gave way to table-driven compilers, which may give way to rule-based compilers as ARPAs efforts towards massive parallelism reach fruition early in the next century. The proposed methodology is entirely consistent with massively-parallel architectures and is in fact one of the few software-development methodologies to lay claim to such.

The reader will no doubt be wondering about the implementation and perhaps question the need for the expert^n-system when a good nomenclature scheme will facilitate easier software retrieval. The only problem with the nomenclature scheme is that it off-loads the need for domain-expertise onto the user. For example, do I want to use the software block that computes the beam stress on turbine blades coated with tungsten carbide using a sputtering process, or that for pure titanium blades? Sounds confusing? Well, now you get an idea of how the expert system can query the user to extract information to make the proper choice of software in accordance with the user's needs. Nevertheless, expert retrieval does not appear to be worth the effort unless say 100,000 lines of code or more (i.e., broken into block modules) is to be retrieved. Otherwise, it would be simpler to create and maintain on-line documented help. This help must include a stub to corresponding software (hierarchical) block translation as one of its divisions. On-line help systems have recently evolved into rule-based decision support and Help-Desk systems (e.g., Bendata's HEAT system).

Conclusion

The design and development of software is no doubt the most complex endeavor yet created by human-kind. The automation of this task has so far proven elusive; yet, some guidelines have emerged as a result of our summer fellowship. First and foremost, the task must be reduced to its salient features. This implies that a 4th or 5th generation language should be used (or at least considered if not available) for the operational domain. Next, non-trivial software systems need to be modeled, using stubs, in an appropriate simulation environment. The environment must provide for the storage and retrieval of block components for reuse. It should also include an intelligent rule-based retrieval mechanism whose rules may be developed using the same hierarchical block-method used to model the software. This requires that the system be self-referential. A capability for self-reference here was demonstrated during the summer of 1995. The repository must also pair software stubs with their full-blown realization. This can be a trivial task where the stubs are mere instances of their generalizations. However, it will be more useful if the stubs can serve to transform their coded representation into an appropriate full-blown instance. Our research results this past summer have shown that the simplest way to accomplish this is to reduce the granularity of the reusable code. Then, and only then, the best way to proceed is to manually tailor the code -- ideally under the tutelage of an expert^n-system in the large. Reuse implies that any redundancy in tasks presented to the user, or knowledge engineer, is continually minimized. That is the key to software automation and the attendant increased productivity for less cost.

Acknowledgment

References

1. Bajpai, A. 1994. "An expert system approach to design of automotive air-conditioning systems," *Artif. Intell. Eng. Design, Anal., and Mftg.*, Winter, vol. 8, no. 1, pp. 1-11.

2. Chaharbaghi, K. & Nugent, E. 1995. "Creativity and competitiveness," *Mftg. Engr.*, Apr., vol. 74, no. 2, pp. 60-62.

3. Darr, T.P. & Birmingham, W.P. 1994. "Automated design for concurrent engineering," *AI Expert*, Oct., pp. 35-42.

4. Friel, P.G., Mayer, R.J., Lockledge, J.C., Smith, G.M., & Shulze, R.C. 1989. "Collsys: A cooling systems design assistant," In H. Schorr and A. Rappaport (eds.), *Innovative Applications of Artificial Intelligence*, Cambridge, MA: The MIT Press.

5. Garrett, J.H. Jr. & Jain, A. 1988. "ENCORE: An object-oriented knowledge-based system for transformer design," *AI EDAM*, vol. 2, no. 2, pp. 123-134.

6. Gelsey, A. 1995. "Automated reasoning about machines," *Artif. Intell.*, vol. 74, no. 1, pp. 1-53.

7. Gorman, M.E., Richards, L.G., Scherer, W.T., & Kagiwada, J.K. 1995. "Teaching invention and design: Multi-disciplinary learning modules," *J. Eng. Edu.*, Apr., vol. 84, no. 2, pp. 175-185.

8. Hoeltzel, D.A., Wei-Hua, C., & Zissimides, J. 1987. "Knowledge representation and planning control in an expert system for the creative design of mechanisms," *AI EDAM*, vol. 1, no. 2, pp. 119-137.

9. Issa, G., Shen, S., & Chew, M.S. 1994. "Using analogical reasoning for mechanism design," *AI Expert*, Jun., pp. 60-69.

10. Jones, P.M., Chu, R.W., & Mitchell, C.M. 1995. "A methodology for human-machine systems research: Knowledge engineering, modeling, and simulation," *IEEE Trans. Syst., Man, Cybern.*, vol. 25, no. 7, Jul., pp. 1025-1038.

11. Jones, P.M. & Mitchell, C.M. 1995. "Human-computer cooperative problem solving: Theory, design, and evaluation of an intelligent associate system," *IEEE Trans. Syst., Man, Cybern.*, vol. 25, no. 7, Jul., pp. 1039-1053.

12. Lutton, L. 1995. "HYPEREX -- A generic expert system to assist architects in the design of routine building types," *Building and Environment*, vol. 30, no. 2, pp. 165-180.

13. Nakashima, Y. & Baba, T. 1989. "OHCS: Hydraulic circuit design assistant," In H. Schorr and A. Rappaport (eds.), *Innovative Applications of Artificial Intelligence*, Cambridge, MA: The MIT Press.

14. Rasmus, D.W. 1995. "Creativity and tools," *PC AI*, vol. 9, no. 4, pp. 23-33.

15. Rubin, S.H. 1991. "Learning in the large: Case-based software systems design," *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Charlottesville, VA, pp. 1833-1838.

16. Rubin, S.H. 1992a. "Case-based learning: A new paradigm for automated knowledge acquisition," *ISA Trans.: Special issue on artificial intelligence and competitive manufacturing*, vol. 31, no. 2, pp. 181-209.

17. Rubin, S.H. 1992b. "Intelligent compilation: Bootstrapping case-based learning," *Heuristics: The J. of Knowledge Eng.*, Spring, vol. 5, no. 1, pp. 13-43.

18. Sandler, B.-Z. 1994. *Computer-aided creativity*, New York, NY: Van Nostrand Reinhold.

19. Savolainen, T. & Cantamessa, M. 1995. "The creative agent in CIM modeling," *Comput. in Indus.*, vol. 25, no. 3, pp. 295-308.

20. Steele, R.L., Richardson, S.A., & Winchell, M.A. 1989. "DesignAdvisor: A knowledge-based integrated circuit design critic," In H. Schorr and A. Rappaport (eds.), *Innovative Applications of Artificial Intelligence*, Cambridge, MA: The MIT Press.

21. Suh, C.-K., Suh, E.-H., & Lee, D.-M. 1995. "Artificial intelligence approaches in model management systems: a survey," *Comput. & Indus. Engr.*, vol. 28, no. 2, pp. 291-299.

22. Utterback, J.M. 1994. *Mastering the dynamics of innovation*, Boston, MA: Harvard Business School Press.

23. VerDuin, W.H. 1992. "The role of integrated AI technologies in product formulation," *ISA Trans.: Special issue on artificial intelligence and competitive manufacturing*, vol. 31, no. 2, pp. 151-157.

24. Wilson, D. 1994. "How to design a product in the year 2000," *Design Engr.*, Sept., pp. 47-50.

** RULE #7:   {Rules are numbered sequentially, starting at zero. The upper limit is given by "maxrule" - 1.}

if   (NowInitializing)

IF
                                                    {Check to see if this rule is to be activated.}
(

        (((NOT) spec i (AND/OR/NOT) spec j ... (AND/OR)
        (NOT) (real/belief i [2] (==, >, <, >=, <=, <>) (real/belief j [2])) ... (AND/OR)
        (Comments)) ... (AND/OR))

)

{
                                                    {Note brace type}
        {Next, the certainty factor, CF, is computed; where, the following i, j, and (k) usually correspond
        to all and only those spec i, spec j, and (spec k), above. x is a real in (1.0, ...), y is a real in (0.0, ...).}

        CF = belief i [2] * belief j [2] * (belief k [2])...; or   {i.e., Select one of these three schemas.}

        CF = (belief i [2] * belief j [2] * (belief k [2])...)/x; or   {to bring CF closer to zero in proportion to x.}

        CF = (belief i [2] * belief j [2] * (belief k [2])... + y)/(y + 1.0);   {to bring CF closer to unity via y.}

        ThisisRule (7);   {current rule number}

}

        else; else if (NowFiring == 7)   {i.e., currently firing this rule}

{

        (if (NOT CallBack))   {This conditional must be included iff the rule includes a
                                            CALLABEL command (see below).}
        {
                (((Message to User) ...   {i.e., PlaySound, Say, ShowQuickTime, ShowSlide, Speak, UserError,
                                            UserPrompt, GetString, GetReal -- to read local vars. Note that the
                                            last two are are based on the system function, UserParameter. }

        e.g.,   Speak ("You can speak before any delay device, such as the following UserError.");

        e.g.,   UserError ("Rule 7: Use block (model) ""Shuttle Demo"" in ""ADA Lib"". "+
                    "The level of certainty is "+certainty [FiredRules]+" percent.");

        e.g.,   if (!UserPrompt ("Rule 7: The chance of a successful launch is "+certainty [FiredRules]+
                    " percent. Do you want to execute the launch?"))
                            Canceled = TRUE;
                    else
                    {
                            ShowSlide ("Butterfly", 10.0, FALSE);   // 10 seconds, do not open dialog

                            // Note: The dialog box will close and open again for the movie:

                            ShowQuickTime ("Launch Movie", 1.0, TRUE); // normal speed, open dialog
                    }

        e.g.,   Say ("Do you understand what I am saying?", 3.5);
                    if (UserPrompt (""))
                            understanding = TRUE;
                    else
                            understanding = FALSE;

        e.g.,   PlaySound ("Raspberry");

        {An example of GetString and GetReal follows. The results of the calls are returned as function
        values. Reals can be converted to integers. If a real number is also an integer, it will be exactly
        converted. Otherwise, the fractional value will be truncated in conversion.}

e.g.,   yourstringvar = GetString ("What would you like on your pizza?", "Nothing");
        {e.g., yourstringvar == "Nothing" by default}

e.g.,   yourrealvar = GetReal ("How many anchovies would you like on your pizza?", 0);
        {e.g., yourrealvar == 0 by default}
        yourinteger = GetReal ("How many anchovies would you like on your pizza?", 0);
        {e.g., if the value returned by the function is 0.99999..., then yourinteger == 0, not 1.}

(Message to File) ...   {Messages may be placed anywhere in the consequent. The "NOT Canceled" predicate is optional. It will prevent rules, canceled at runtime, from being written to the specified file path. Note that a rule comment is always written, to the file path, iff the rule calls any label (i.e., including itself) and is canceled, by the user, at runtime.}

e.g.,   if (OutToFile AND NOT Canceled)
        {
            FileWrite (filenum, "Rule 7: Use block (model) ""Shuttle Demo"" in ""ADA "+
            "Lib."", "", TRUE);
            FileWrite (filenum, "   The level of certainty was "+
            certainty [FiredRules]+" percent.", "", TRUE);
        }
}

(Comments)) ...)

{The following two assignment functions may be placed anywhere in the consequent.}
(spec i = FALSE/TRUE;) ...   {Can set to another spec j, or to any boolean function.}
(belief i [2] = real value in [0.0, 1.0]) ...   {Can set to another belief j [2], or to any real function.}

(CALLABEL (label))   {Note: Only one active call per rule because different calling sequences can produce different results. This is essentially sequential programming and would serve to prevent modular independence in the scripting of the rules. Modular independence is critical to minimizing the manual effort required in debugging and maintaining a knowledge base. The use of CALLABEL is optional, but must follow any messages, which can set, "Canceled = TRUE". CALLABEL must also preceed any SET commands, if they are used. If label specifies the currently executing block, then it will start running this block from the top. This is useful if any spec or belief, in the currently executing block, has been changed (e.g., by call-back). Do not specify the currently executing block if the current rule is found to lie on a cyclic path -- in order to break it. "TRUE" is a reserved word and should not be used as a label (or on return if CALLABEL is used). A block may call its caller -- if care is exercised to prevent infinite regression. CALLABEL may be the result of a conditional expression; although, even though the results may be mutually exclusive, this is not consistent with good knowledge engineering practice. Don't forget to use the "if (NOT CallBack)" conditional in conjunction with CALLABEL (see above). If a block has just called itself, then the bottom of its icon will have the same label in blue and black lettering. The blue lettering will be replaced with the word, "Call-Back", lettered in orange, on call-back. Manual Back is lettered in purple. The master referee is always designated on the top of the appropriate icon and is lettered in black. Use these labels as an aid in debugging. Note that blocks cannot modify themselves using callabel and what follows. You may however force modifications using, spec i = true or false and belief i [2] = real value.}

((SET_NEXT_SP_BE) ...   {See below}
(SET_RETURN_SP) ...   {See below}
(SET_RETURN_BE_R) ...   {See below}

(Comments) ...)   {Note: The SET commands are optional and may only be used in conjunction with CALLABEL, which must preceed it in the script. They may be sensitive to order; but, a NEXT command will always be executed before the other two RETURN types.}

}

Firenext ();   {Fire the next rule.}
return (FALSE/TRUE);   {This must be the last statement of every rule. Use FALSE for a normal con-
                                 tinuation. Use TRUE for immediate return to the calling block, if any. The
                                 calling rule will be immediately re-entered on return. TRUE may not be used
                                 to return from any rule, which uses CALLABEL. The rule base may be man-
                                 ually re-run, since specs and beliefs may have been altered.}

The SET_NEXT_SP_BE function follows.

SET_NEXT_SP_BE (i, COMPLEMENT/FALSE/TRUE/OK, r);
        will set spec i and/or belief i [2], in the next block, as follows.

The First arg, i, is the integer specification or belief number to set in the specified block.

For the Second arg:
        COMPLEMENT will set the spec to the complement of this one.
        FALSE or TRUE will set the spec to this value.
        OK will not change it.
        The second arg can be any spec j. The compelement, (NOT spec j) may also be used. More
        complex expressions may be built using AND, OR (e.g., NOT spec 1 AND NOT spec 2 OR
        NOT spec 3...), as well as boolean functions. Note that all of the NOTs are done before any-
        thing else (i.e., except parenthetical expressions and negation signs).

For the Third arg:
        If (..., -1.0), then belief j [2] = belief i / -r; to bring belief j [2] closer to zero in proportion to r.
        If [-1.0, 0.0), then set belief j [2] to 1 + belief i.
        If [NA], then the belief j [2] will not be changed.
        If (0.0, 1.0], then set belief j [2] to belief i.
        If (1.0, ...), then belief j [2] = (belief i + r)/(r + 1); to bring belief j [2] closer to unity in proportion to r.
        The third arg can be any belief k [2]. The complement, (1.0 - r) may also be used. More complex
        expressions may be built using standard operators and functions. Care should be taken, however,
        that the result is bounded as defined above. Note that any computed belief of zero will automatically
        uncheck the corresponding specification prior to execution.

e.g., SET_NEXT_SP_BE (1, TRUE, 0.50);   spec1 is true; belief1 [2] is 0.50
e.g., SET_NEXT_SP_BE (2, FALSE, 1.0);   spec2 is false; belief2 [2] is 1.0

The SET_RETURN_SP and SET_RETURN_BE_R functions follow.

SET_RETURN_SP (i, +- j);
        will set this spec i to the returned spec j. The negative sign is interpreted as NOT.

SET_RETURN_BE_R (i, j, r);
        will set this belief i to the returned probability [j] as varied by the real r.

        If (..., -1.0), then belief i [2] = returned probability [j] / -r; to bring belief i [2] closer to zero in proportion to r.
        If [-1.0, 0.0), then set belief i [2] to the returned probability [j] * (1 + r).
        If [NA], then belief i [2] is simply assigned the returned probability [j].
        If (0.0, 1.0], then set belief i [2] to the returned probability [j] * r.
        If (1.0, ...), then belief i [2] = (returned probability [j] + r)/(r + 1); to bring belief i [2] closer to unity via r.
        The third arg to SET_RETURN_BE_R can be any real number -- including belief k [2], and the
        returned probability [k]. The complement can be created using the argument, (1.0 - r). More complex
        expressions may be built using standard operators and functions. Care should be taken, however,
        that the result is bounded as defined above. Note that any computed belief of zero will automatically
        uncheck the corresponding specification prior to execution.

e.g., SET_RETURN_SP (0, 1);   // Set this spec 0 to the returned specification [1].
e.g., SET_RETURN_BE_R (0, 1, NA);   // Sets this belief 0 [2] to the returned probability [1].

A FLEXIBLE ARCHITECTURE FOR COMMUNICATION SYSTEMS (FACS): SOFTWARE AM RADIO

John L. Schmalzel
Professor of Electrical Engineering
School of Engineering

Rowan College of New Jersey
201 Mullica Hill Road
Glassboro, NJ 08028-1701

A FLEXIBLE ARCHITECTURE FOR COMMUNICATION SYSTEMS (FACS): SOFTWARE AM RADIO

John L. Schmalzel
Professor of Electrical Engineering
School of Engineering
Rowan College of New Jersey

Abstract

Interest in *software radio* is driven by the search for universal communication system architectures consisting of minimum radio frequency (rf) front-end hardware, which perform as much radio functionality as possible in software. An investigation into flexible software radio architectures for amplitude modulated (AM) broadcast spectrum (540-1680 kHz) signals was conducted. Using a variety of commercial data acquisition platforms and personal computer based software, the requirements for an AM software radio have been defined and are described, allowing the further development of a FACS platform for testing software radio concepts.

The elements of the AM software radio include: (i) A front-end with gain to raise detected signals to a magnitude within the conversion range of the analog-to-digital (A/D) converter, and an anti-aliasing low-pass filter (LPF). A nominal 1 μV rf signal requires a gain of 1E06 (60 dB) to provide a 1 V magnitude compatible with many high-speed A/D converters. (ii) An A/D converter chosen to sample at a rate sufficiently higher than the bandwidth of the AM broadcast band to avoid aliases and to allow the use of a low-order LPF. (iii) A digital signal processor (DSP) or other general purpose computer to perform tuning using a bandpass filter (BPF) and demodulation with a lossy peak detector. (iv) Final audio output formatted as a *.wav* file for compatibility with standard multimedia sound (*Soundblaster*) cards.

A FLEXIBLE ARCHITECTURE FOR COMMUNICATION SYSTEMS INVESTIGATION: SOFTWARE AM RADIO

John L. Schmalzel

1. Introduction

Communication systems are experiencing rapid evolution as the demand for increased services by an expanding market base places pressure on available spectrum. Communication technologies are also key global competitive issues. This has spurred the development of a large number of international standards--many of which are incompatible. Designing world products that are capable of seamless interface to diverse communication standards is challenging; it favors those architectures that have the most software flexibility. In the limit, the most flexible of communication architectures is termed *Software Radio* and connotes reliance on communication algorithms hosted by high performance DSP architectures in place of traditional communication hardware subsystem blocks.

There is a broad--and expanding--community interested in communication systems architectures [1] to meet a growing number of voice, data, video, and multimedia telecommunication requirements. For example, major growth markets in cellular telephony, personal communication services (PCS), and mixed services on Community Antenna Television (CATV) systems, are tangible evidence of this trend. The availability of a flexible test and measurement architecture would facilitate investigations of communication systems. For example, it would allow fundamental investigations into coding or modulation performance, and support applications requiring flexibility to understand and characterize novel sources like those frequently encountered in Electronic Warfare (EW) scenarios.

The communication system model chosen for the FACS is standard broadcast AM. This is a low frequency source spectrum (540 kHz - 1.68 MHz) compatible with much off-the-shelf instrumentation as well as component-level technologies.

A generalized block diagram of a software radio is shown in Fig. 1. While the analog hardware blocks of a conventional radio offer high bandwidth and very low cost, they do not offer the flexibility required to service extended bandwidths, modulation techniques, and coding formats. Of course, development of software radio systems would ultimately lower the unit price of this technology, working steadily down the price/performance curve until they favorably compete with representative communication systems in each class. While it is difficult--in 1995 to envision software radios that could compete with traditional broadcast AM or FM receivers costing less than ten dollars, it is less difficult to forecast near-term competition with cellular and other packet radio technologies and other specialized communication system applications such as encountered in military and secure environments.

Elements of software radios have been around for a long time. Any communication related application that includes a software portion is arguably a form of software radio. Currently, many data communication technologies such as modems and local area network (LAN) systems have matured to the point where they represent a nearly complete form of software radio. For example, application data for a first-generation DSP chip, the TMS32010 [3] includes a block
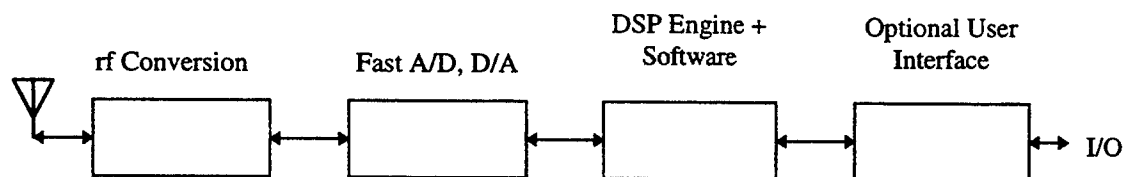

Figure 1 Block diagram of a software radio. (Adapted from [2], Fig. 2.)

diagram of a digital modem (shown in Fig. 2) that is a good fit to the generalized model.

Refinement of algorithms and hardware enables single (or few) chip solutions that rely on firmware to do much of the signal processing required to execute the particular algorithm. Similarly, multimedia standards have spawned new generations of Very Large Scale Integrated Circuits (VLSI) coprocessors that offload the specialized coding/decoding algorithms such as
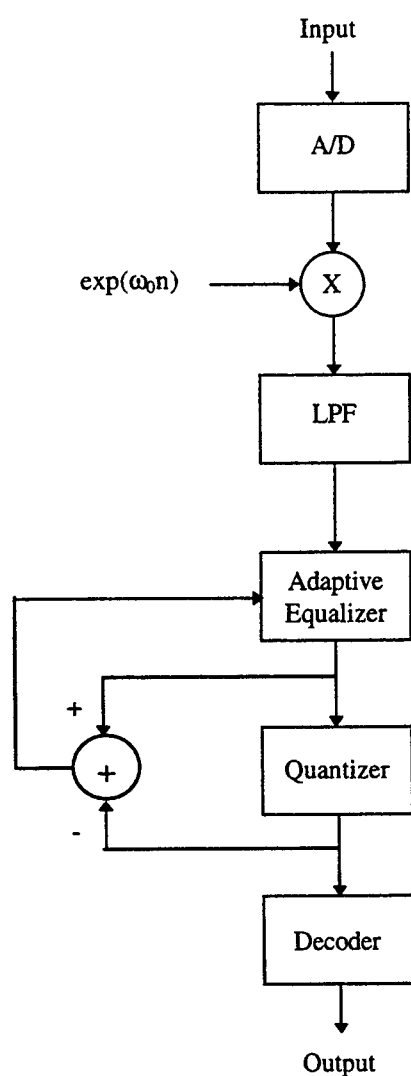
Input

A/D

$\exp(\omega_0 n)$ ——→ X

LPF

Adaptive Equalizer

Quantizer

Decoder

Output

Figure 2 Block diagram of an early DSP-based modem that exhibits features of the generalized software radio. The LPF, Adaptive Equalizer, Quantizer, and Decoder were all implemented using a DSP chip. (Adapted from [3], Fig. 8-18.)

48-5

Motion Pictures Experts Group (MPEG) that must be executed in real time. The same can be said for CD-ROM (data and music) and many others. Thus, the notion of a "software radio" appears to be well established; the creation of a new moniker may serve to refocus attention and provide a technology viewpoint needed to create a paradigm shift in how communication systems are viewed and organized in the future.

Broadly, a flexible architecture for communication systems (FACS) should support both receiver demodulation and decoding functions, and transmitter modulation and coding functions. For the entry-level FACS, the AM receiver function was selected for the first emphasis since broadcast transmission sites are well represented in most areas. Emphasis on the receiver side is also compatible with parallel investigations into cellular telephony, global positioning system (GPS) satellite systems [4], and other aspects of broadcast AM and FM radio.

Desirable features of the Phase I FACS include:

(i) Minimum explicit communication hardware; i.e., configuration of a minimum front end (gain, selectivity) stage with only sufficient down conversion--if any--to provide the bandwidth reduction needed for the following analog-to-digital (A/D) conversion stage.

(ii) An analog-to-digital (A/D) conversion subsystem to acquire the raw--or near baseband--communication signals for further processing, using simple data structures for interfacing w/ a host personal computer (PC).

(iii) Provision for flexible signal processing of the received signal block using traditional or nontraditional digital signal processing algorithms for filtering, demodulation, error correction, and other steps needed to produce the final output. For example, filtering using a variety of filter types (LPF, BPF, Comb), and application of transform techniques such as discrete cosine (DCT),

fast Fourier (FFT), wavelet transforms, etc. [5,6]. The DSP functions could also be performed by including DSP hardware hosted by the PC.

(iv) Identification of key system bottlenecks where specialized hardware and algorithms can significantly improve certain aspects of overall FACS performance. For example, including high-speed hardware digital filtering structures [7-9] could extend the bandwidth of a system otherwise limited by the maximum sampling rate, F_S of the A/D or the maximum computational bandwidth of the DSP chip.

2. Amplitude Modulation: Principles and FACS Counterparts

AM theory and technology are historic and well developed [10] with low-cost hardware systems available off the shelf. In the material that follows, key AM concepts are presented along with their FACS/software counterparts.

2.1 Generation of AM.

2.1.1. General. In AM radio, a carrier, ω_c, is modulated to obtain a modulated carrier; this modulated carrier is what conveys information to the receiver:

$$x_c(t) = A_c[1 + \mu x(t)] \cos\omega_c t \tag{2.1}$$

where A_c is the unmodulated carrier amplitude, μ is the modulation index (positive, $<= 1.0$), and $x(t)$ is the modulating signal--e.g., speech. Expansion of (2.1) yields

$$x_c(t) = A_c\cos\omega_c t + A_c\mu x(t)\cos\omega_c t \tag{2.2}$$

which more clearly demonstrates that there are two major features transmitted by AM: (i) the carrier, and (ii) a signal-modulated carrier. Figure 3 shows the relationship among the signal components. The 1-sided Fourier Transform also shows where the signal components are distributed.
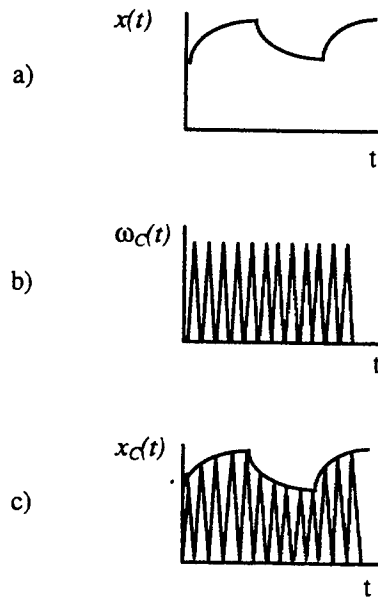
Figure 3 AM modulation. In (a), the signal $x(t)$ is shown. It is multiplied by a carrier, $\omega_c(t)$, shown in (b), to produce the modulated signal, $x_c(t)$ shown in (c).

$$X_c(f) = (A_c/2)[\delta(f - f_c) + \mu X(f - f_c)] \tag{2.3}$$

2.1.2. FACS. Generation of AM signals requires only that a carrier be synthesized and then the multiplied by the input signal waveform.

2.2 Front End.

2.2.1 An AM receiver must have sufficient rf bandwidth to process a range of frequencies equal to the highest carrier plus the highest signal frequency. In typical AM it is presumed that $f_c \gg f$, so the required receiver bandwidth is approximately f_c. In addition, the front end must provide sufficient gain and also provide channel tuning.

2.2.2 FACS gain and A/D conversion. The objective of the FACS front end is sufficient processing to prepare the rf signals for A/D conversion. This requires two essential elements: gain and anti-alias, low-pass filtering. Gain is a function of the full-scale input voltage range of the A/D and the number of bits of conversion that it supports. For example, The AD 7620 [11] is

a 6-bit, 80 Msps converter with a 1 Vpp input range. An input signal of 1 μV requires a gain of 1E06 (60 dB power gain) to raise it to the full-scale input voltage range. An MP 7684 [12] is an 8-bit, 20 Mbps flash converter with a 5-V full scale, requiring a gain of 5E06 (67 dB power gain) to raise a 1 μV signal to full scale. Another factor that affects total gain is the final number of bits required out of the conversion. For example, if only 6-bits are required, an 8-bit A/D could be used with correspondingly less gain; in this case, 1/4 less ($2^{-(8-6)}$). Further reduction in gain requirement could be achieved with A/Ds having more bits; for example, the HP E1430 contains a 23-bit, 10 Msps converter, which could imply a reduction in gain of $2^{-(23-6)}$, implying that only a gain of 10 would be needed! Of course, noise and actual amplitude considerations (some AM signals are orders of magnitude above 1 μV) will affect the final combination employed.

2.2.3 FACS anti-alias low-pass filter. A simple low-pass filter is needed before the A/D to minimize aliases. For example, assuming a worst-case conversion rate of 10 Msps, and a requirement of at least 40 dB attenuation at the Nyquist frequency of 5 MHz, a relatively low-order filter is required. Modular filters are available; e.g., an acceptable Lark [13] LDP series filter for this application is a LDP 2 - 3 AA (Specifies a standard 50 Ω, 2 MHz cutoff, 3-section filter with SMA jacks for input and output connectors). If desired, a custom filter could be designed; for example, a Butterworth LPF of order n = 3 (where $n = \log[(10^{Amin/10}-1)/(10^{Amax/10}-1)]/2\log(f_S/f_p)$) would provide at least 40 dB (A_{min}) of stopband attenuation at f_S = 10 Msps, with no greater than 3 dB (A_{max}) of passband attenuation at the passband edge, f_P = 1.7 MHz.

2.2.4 FACS Channel tuning. Channel tuning must be accomplished using a tunable digital bandpass filter (BPF). Briefly, one method of digital filter design [5] begins with a prototype s-domain transfer function, H(s), applies pre-warping of the desired analog frequencies, then uses the bilinear transform to obtain the H(z) which is then inverse transformed to get the h(n) used to

48-9

implement the filter in software. The design steps are illustrated for an example station centered on 540 kHz. Beginning with,

$$\omega_{A1} = \tan(\pi F_L/F_S), \quad \omega_{A2} = \tan(\pi F_U/F_S), \quad W = \omega_{A2} - \omega_{A1}, \quad \text{and} \quad \omega_0^2 = \omega_{A1}\omega_{A2}, \tag{2.4}$$

these are substituted into the prototype H(s),

$$H(s) = Ws/(s^2 + Ws + \omega_0^2) \tag{2.5}$$

where W is the bandwidth, ω_0^2 is the natural frequency (squared); and F_L and F_U are the lower and upper passband edges, respectively. The bilinear transform, $s = (z-1)/(z+1)$, is applied to (2.5) to obtain,

$$H(z) = W(1 - z^{-2})/ [(1 + W + \omega_0^2) + z^{-1}(-2 + 2\omega_0^2) + z^{-2}(1 - W + \omega_0^2)] \tag{2.6}$$

For the channel at 540 kHz, $F_L = 535$ kHz, $F_U = 545$ kHz; with a sampling frequency, F_S,, of 10 Msps, evaluation of the terms in (2.4) and substitution into (2.6) yields,

$$H(z) = 0.00321(1 - z^{-2})/(1.0325 - 1.9414z^{-1} + 1.0261z^{-2}) \tag{2.7}$$

which in turn generates the required difference equation for the filter by applying the inverse Z-Transform, and noting that $H(z) = Y(z)/X(z)$, $Z^{-1}\{Y(z)\} = y(n)$, $Z^{-1}\{z^{-1}Y(z)\} = y(n-1)$, etc., yields

$$y(n) = (W/a)[x(n) - x(n-1)] - (b/a) y(n-1) - (c/a) y(n-2)$$

$$(2.8)$$

where $a = (1 + W + \omega_0^2)$, $b = (-2 + 2\omega_0^2)$, and $c = (1 - W + \omega_0^2)$.

Channel tuning consists of accepting the desired station from the user and looking up the pre-computed filter coefficients and then using the coefficients in the fixed filter topology. Each sample requires three floating-point (FLP) additions (subtractions), and three FLP multiplications (assuming that $1/a$ is pre-computed and combined with the other scalar coefficients, W, b, and c. Table 1 contains example coefficients.

Table 1 Filter coefficients for selection of broadcast AM stations at band edges.

Center kHz	F_L	F_U	ω_{A1}	ω_{A2}	W	ω_0^2	a	b	c
540	535	545	0.1697	0.1729	0.00321	0.0293	1.0325	-1.9414	1.0261
1680	1675	1685	0.5808	0.5851	0.0043	0.3398	1.3441	-1.3204	1.3355

2.3 AM Demodulation.

2.3.1 General. The AM signal, $x(t)$, must be recovered from the composite signal, $x_C(t)$. This can be accomplished in two fundamental ways: (i) synchronous detection, in which the input signal is multiplied by a local oscillator, and (ii) peak (or envelope) detection. Synchronous detection is a more general scheme that is compatible with several other linear modulation methods such as double side band (DSB), single side band (SSB), etc. In contrast, peak detection is a simple technique that avoids generation of a product by sampling the peaks of the modulated carrier to obtain an envelope estimate which is an estimate of $x(t)$.

2.3.1.1 Synchronous detection. The modulated input signal, $x_C(t)$, available at the receiver, is multiplied by a sinusoid, $A_{LO}\cos\omega_C t$, generated by a local oscillator (LO) to yield sum and difference frequencies to form the detected signal, $x_D(t)$. If the LO is identically equal to the carrier frequency, the sum and difference frequencies are $2\omega_C$ and 0 respectively. That is, $A_{LO}\cos\omega_C t$ is multiplied with (2.1) to obtain,

$$x_D(t) = (A_c A_{LO}/2)\{[1 + \mu x(t)] + [1 + \mu x(t)]\cos 2\omega_C t\} \qquad (2.7)$$

If this signal, $x_D(t)$, is then low-pass filtered as suggested by Fig. 4, using a cutoff frequency approximately equal to the signal bandwidth, B, the result is a dc term with a scaled signal term:

$$y_D(t) = (A_C A_{LO}/2)[1 + \mu x(t)] \hspace{3cm} (2.8)$$

In a typical AM receiver, the dc term is blocked using simple ac coupling and the remaining signal

term is then amplified sufficiently to achieve the desired audio volume.

2.3.1.2 Peak detection. The simplest AM demodulation technique takes

advantage of the fact that the peak value of the raw receiver signal is equal to the envelope of the

signal term as was shown previously in Fig. 3(c). The electronics required for a peak detector and

corresponding high-pass filter to block the dc component can be implemented with the simple

diode network shown in Fig. 5.

2.3.2 FACS demodulation. The simple peak detection approach is adopted.

Algorithmically, peak detection can be achieved using the pseudocode segment below:

```
procedure peak_detect
     while data stream is present
          while no zero crossing              ;find a zero crossing
          end while
               while |value| > peak           ;this finds the peak value
                    peak = |value|             ;assign this peak value to the
                         demodulated_sample = peak   ; output data sequence
               end while
     end while
```
The peak detector loss function (R1 in Fig. 5) is trivial to implement in that the peak

detection algorithm is essentially reset on each half cycle of the carrier; memory of previous peak

values is only present in the output data sequence. Additional filtering can be applied at this
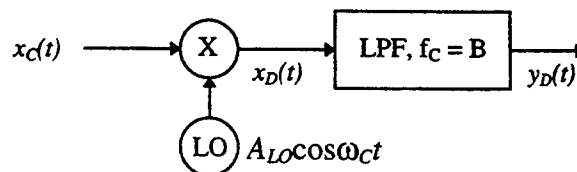

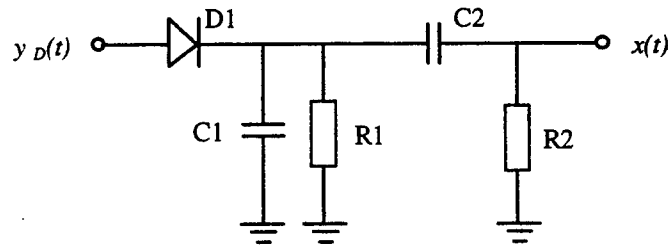**Figure 4** Block diagram of synchronous AM detection.

Figure 5 Fundamental peak detector. D1 half-wave rectifies the input signal, $y_D(t)$. The low-pass filter formed by C1-R1 determines the upper bandwidth of the detector stage; the high-pass filter formed by C2-R2 determines the lower passband and is used to block the dc term of (2.8).

point--e.g., a moving average filter (MAF) to provide immunity from impulsive noise which is characteristic of AM. Another advantage of maintaining a moving average is that it is a simple way of achieving ac coupling: the average (dc) value estimated by the MAF is simply subtracted from the output data sequence to obtain a zero average value as shown in Fig. 6. Efficiency can be achieved by not forming a new N-sum, but rather subtracting the oldest and adding the newest:

```
procedure MAV
        while data present
                sum = sum - x(0) + x(N)
                ave = sum/N
        end while
```

2.4 Audio Output.

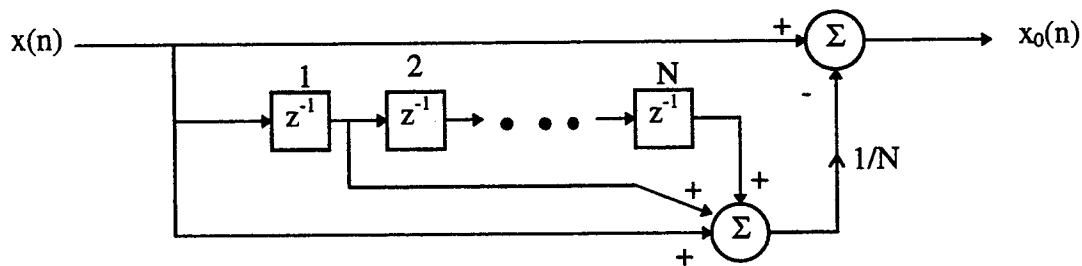2.4.1 General. The demodulated signal is amplified so as to drive a speaker system.

Figure 6 Level shifting (zero restore) output sequence by subtracting moving average.

48-13

2.4.2 FACS audio output. A direct approach would use a digital-to-analog (D/A) converter to drive a power amplifier stage connected to speakers. Alternatively, widely available digital audio hardware can be employed. PCs with multimedia capability include 16-bit sound cards (for example, *Soundblaster*, by Creative Labs). PCM data files stored as *.wav* files are presented to playback software. A mono channel consists of a contiguous data stream; stereo channel values are interleaved in the data stream.

3. Investigation of Elements of the Phase I FACS

<u>3.1 Equipment used for investigation</u>. Several items of equipment were available to facilitate investigation of development issues for the Phase I FACS:

(i) Digital Storage Oscilloscope (DSO). A Tektronix Model TDS 684A was available at the sponsoring laboratory; it provides up to 1.5 Gsps acquisition to 6/8-bit resolution and is configurable via IEEE-488 to both define acquisition parameters as well as obtain data points. The DSO's command language is closely related to *Standard Commands for Programmable Instrumentation* (SCPI) [14,15]. A similar unit from Hewlett-Packard was also used in the university lab (HP 75000 System: E1421A, Mainframe; E1406A, Commander; E1430, 10 Mbps, 23-bit A/D; and E1488A, 10 MB memory module). The advantage of testing with such instruments is the availability of gain and bandwidth controls, but memory is limited and it is not possible to sustain real-time data streams.

(ii) Data acquisition, analysis, and display. Two software tools were used for this effort. Lab-Windows (National Instruments) and HP-VEE (Hewlett-Packard) provide integrated data acquisition, analysis, and display software for PCs with IEEE-488 interfaces. They allow development of routines for DSO control and data acquisition. (The Lab Windows program developed to control the TDS 684A is available [16]). HP VEE is similar to Lab-View (National

Instruments) in that icons are used to define signal flow and signal processing steps--this offers

distinct advantages for generalized architecture development projects such as the FACS.

4. FACS Implementation Issues

4.1 General. The model FACS software AM radio consists of the elements shown in Fig. 7 and
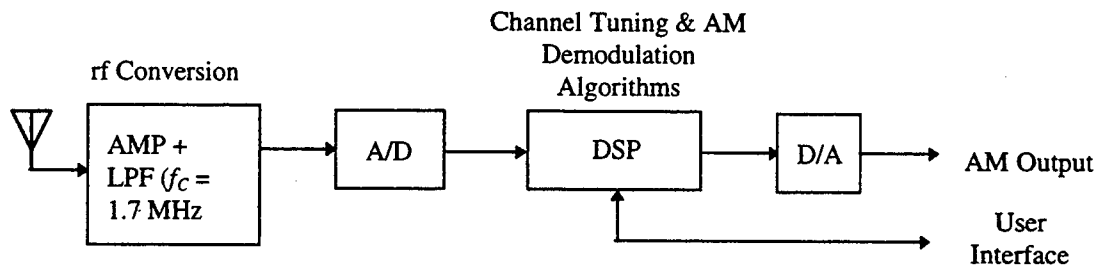

Figure 7 Block diagram of the software AM radio.

follows directly from the general software radio model of Fig. 1. It supports the signal processing

steps described previously.

4.2 DSP Resource Estimation. A development objective for the FACS is the eventual

implementation of substantial portions of the architecture in PC-hosted or standalone DSP

engine(s) with supporting front-end and conversion (A/D, D/A) hardware. Moving the radio

functions into DSP hardware and software requires that reasonable estimates of processing

bandwidth requirements be made. As a first step, general estimations from the literature are used

and then compared with the computational load estimates derived from the AM architecture

developed above.

Mitola [2] uses a general estimation formula to determine DSP processor demand, D:

$$D = D_{if} + N(D_{bb} + D_{bs} + D_s) + D_0 \qquad (4.1)$$

D_{if} is the IF processing demand and is proportional to acquisition bandwidth and the complexity

of the filters required to isolate the service band--e.g., AM--and the subscriber channel within the

service band. For AM, the subscriber band would correspond to the frequency assignment for a broadcast station. The number of maximum simultaneous subscribers is N; in the case of broadcast AM radio, $N = 1$. Demodulation complexity and the bandwidth of the subscriber channel combine to form D_{bb} which is the baseband demand. D_{bs} is bit stream demand and is proportional to data rate and forward error control and other support algorithms. D_s is the source segment demand and depends on the complexity of any required interfaces to the public switched telephone network (PSTN). A final overhead demand factor, D_o, captures miscellaneous and overhead demand factors. Table 2 summarizes demand estimates for FM and AM from [2].

Table 2 Demand estimates for FM and AM.

Appl.	rf, fc	Wa	IF, Wi	Channel Code	Baseband Wi	Bitstream States	Mux	Priv.	Source
FM mobile	VHF	30 kHz	30 kHz	FM	4 kHz	Infinite	PTT	No	Compand
AM	LF	10 kHz	455 kHz	AM	4 kHz	N/A	Cont	N	Linear

Adapting (4.1) to the AM case reduces to a simple sum of three terms:

$$D = D_{if} + D_{bb} + D_0 \tag{4.2}$$

In the material that follows, summary demand estimates are developed and are compiled in Table 3.

4.2.1 D_{if}. The D_{if} term corresponds to the filtering operations required for channel selectivity outlined previously in 2.2 where a general, second-order digital BPF section was developed. Each sample requires three floating-point multiplications and three floating-point additions (subtractions) for each second-order section. Each additional 2nd-order BPF section added to improve channel selectivity would require an addition three multiplications and three additions.

48-16

4.2.2 D_{bb}. Demodulation of the data stream output from the BPF stage consists of the peak detection algorithm described in 2.3. The logic portion of the algorithm can be implemented using conditional branch instructions: one for the zero-crossing detection, and once the zero-crossing is found, a conditional branch would be used to compare current value to the temporary peak. Maintaining the moving average filter (MAF) for the output sequence requires two floating point additions (subtraction) and one floating-point multiplication on a per carrier cycle basis; however, if a sequence length equal to a power of 2 is used, the scaling ($1/N$) operation could be reduced to a barrel shift operation. Note that the MAF structure may also be required for a decimation filter and/or as a pre-smoothing filter in advance of the peak detection. A similar computational burden is presented by the added MAF.

4.2.3 D_o. The D_o term includes all overhead processing steps which include any processing of the output data stream to produce the final audio signal, the overhead involved with A/D and D/A conversion, and miscellaneous overhead functions such as user interface, buffer management, etc. For output processing, the only function that may be required here is a scaling operation which would reduce to a single floating-point multiplication operation, or could also be a barrel shift if integer powers of 2 are used. A/D conversion nominally requires a read and store operation at the arriving data rate. No overhead is budgeted for A/D control such as initiating conversion operations or checking for busy/done status because it is assumed to be free-running and/or is synchronized to the system. D/A overhead consists of a simple write operation occurring once each carrier half-cycle. User interface, buffer management, and other overhead operations should occupy a small fraction of the total processor time; budgeting 1 operation per data sample provides a comfortable safety margin.

4.3 DSP Performance Requirement Summary.

Table 3 Summary of processor operation requirements.

Function	FLP Mult/Div $\times F_S$	FLP Add/Sub $\times F_S$	Fixed-Point Operations $\times F_S$
D_{if} Channel Selection (Assume two BPF sections)	6	6	2 (Read, Store)
D_{bb} Demodulation	0.01	0.02	2 (Conditional Branch)
D_o Overhead	0.01	0.02	3 (Read, Load, Misc.)
Totals	6.02	6.04	7

Table 3 summarizes the likely number of operations required as a function of the sampling rate, F_S. This is important because it clearly demonstrates the interdependency of overall processing burden on the sampling rate. Another important piece of information shown in the table is the dramatic processing bandwidth reduction once the channel section is completed. This is a well-known finding and further motivates the development of special-purpose, high-speed hardware digital filters associated with the front-end functions. It is also important to note that many DSP chips are currently available (Motorola DSP56000 family, Texas Instruments TMS 3X0 family, etc.) which all provide hybrid instructions that perform a multiply-and-accumulate in one instruction cycle.

5. Recommendations for Future Work

The preliminary analysis and investigations performed for this study show the feasibility of configuring a flexible architecture for empirical and theoretical investigations with communication systems and to extending the architecture to DSP hardware. Such a FACS would be a useful tool for developers and investigators and should be pursued as a follow-on development project.

8. Acknowledgments

9. References

[1]   J. Mitola, (Ed.) "Software Radios," *IEEE Communications Magazine*, in Special Issue on Software Radio, May 1995, pp. 4-5.

[2]   J. Mitola, "The software radio architecture," *IEEE Comm. Mag.*, May 1985, pp. 26-38.

[3]   R. Schafer, *et al.* "Digital signal processing," in *TMS320 User's Guide*, Texas Instruments Inc., Richardson, TX, (214) 680-5082, 1985.

[4]   J.J. Spilker, Jr., "GPS signal structure and performance characteristics," *Navigation*, Vol. 25, July 1979, pp. 29-54.

[5]   N. Ahmed and T. Natarajan, *Discrete-Time Signals and Systems*, Reston: Reston, VA, 1983.

[6]   A.V. Oppenheimer and R. Schafer, *Digital Signal Processing*, Prentice-Hall: Englewood Cliffs, NJ, 1975.

[7]   A. Peled and B. Liu, *Digital Signal Processing*, J. Wiley & Sons: New York, 1976.

[8]   J. Schmalzel, D. Hein, and N. Ahmed, "Some pedagogical considerations of digital filter hardware implementation," *IEEE Circuits and Systems Magazine*, 2:1, 1980, pp. 4-13.

[9]   A. Croisier, et al, "Digital filter for PCM encoded signal," US Patent 3,777,130, Dec. 3, 1973.

[10]   A.B. Carlson, *Communication Systems: An Introduction to Signals and Noise in Communication Systems*, Third Ed., McGraw-Hill: New York, 1986.

[11]   Analog Devices, One Technology Way, P.O. Box 9106, Norwood, MA 02062-9106, (617) 329-4700.

[12]   Micro Power Systems, CA.

[13]   Lark Engineering Co., 27151 Calle Delgado, San Juan Capistrano, CA 92675, (714) 240-1233, E-mail: larkeng@larkeng.com.

[14]   (Anonymous), *SCPI-1995*, The SCPI Consortium, 8380 Hercules Dr., Suite P3, La Mesa, CA, 91942, (619) 697-8790.

[15]   V.T. Randal, *Automated Parser Design for Programmable Instruments*, M.S.E.E. Thesis, The University of Texas at San Antonio, 1995.

[16]   Interface program, <TDS15KV1.bas>, available by writing to the author, Department of Electrical Engineering, School of Engineering, Rowan College of NJ, 201 Mullica Hill Road, Glassboro, NJ 08028-1701, (609) 256-4629, E-mail: schmalzel@mars.rowan.edu.

THE ROLE OF SULFUR COMPOUNDS IN FUELS ON THE FORMATION OF ENGINE DEPOSITS

William D. Schulz
Professor
Department of Chemistry

Eastern Kentucky University
Moore Building, #337
Richmond, KY 40475-3124

THE ROLE OF SULFUR COMPOUNDS IN FUELS ON THE FORMATION OF ENGINE DEPOSITS

William D. Schulz
Professor
Department of Chemistry
Eastern Kentucky University

Abstract

The effect of organo-sulfur containing compounds on deposit formation by thermally stressed fuels was studies. Thiols and thiophenols were found to greatly increase the formation of solid, insoluble deposits. Sulfides, disulfides and thiophenes were fount to have little effect on the formation of deposits. Fuels were thermally stressed in air. Polar oxidation products were extracted and analyzed. Deposits were analyzed by sequential thermal desorbtion pyrolysis in a gas chromatograph-mass spectrometer. A hydrotreated, thermally stable fuel that does not produce insoluble deposits when thermally stressed, did produce insoluble deposits when doped with 300 ppm of thiophenol.

These deposits were completely unlike deposits from the undoped fuel in that they were highly phenolic. Undoped fuel does not form insoluble deposits and the gums formed by such fuels contain aliphatic alcohols and carbonyl compounds. The results of the study emphasize a dichotomy of mechanisms in deposit formation. A mechanism of arylhydroperoxide formation, decomposition to phenols and oxidative phenolic coupling as the route of deposit formation by straight-run fuels containing significant amounts of sulfur and substituted aromatic compound is strongly suggested.

THE ROLE OF SULFUR COMPOUNDS
IN FUELS ON THE FORMATION
OF ENGINE DEPOSITS

William D. Schulz

Introduction

Oxidative deterioration of hydrocarbon based fuels and lubricants results in the formation of solid deposits under storage and operating conditions. Deposit formation is undesirable in any mechanical application and especially so in aircraft systems where deposits can foul filters, injector nozzles or heat exchangers to cause aircraft malfunction. Fuel stability to thermal oxidative deposit formation is of increasing concern for advanced aircraft designs that utilize fuel recirculation for cooling and have higher operating temperatures. Fuels for such aircraft will have to be stable to deposit formation at higher temperatures and higher heat loads for longer periods of time. Development of economical and thermally stable fuels requires understanding of fundamental deposit formation mechanisms.

Until recently, the general assumption has been that "gums" (highly polar, acetone soluble deposits) are formed by hydroperoxide decomposition to alcohols, aldehydes, ketones and carboxylic acids, and these are intermediates in the formation of "solids" (acetone insoluble deposits). However, recent work[1-3] has led us to believe that solids can be formed essentially without concurrent gum formation and that such solid formation is dependent upon fuel composition.

Recent investigation indicate a basic division in mechanisms of fuel deposition. Features of the two different observed pathways are distinct and each has been observed for several fuels. The first, oxidative deposition, is typical of model compounds and hydrotreated fuels. In our flask test (static, atmospheric, flowing oxygen, 175 C) these fuels oxidize rapidly, forming a second liquid phase and deposits that are essentially all acetone soluble gums. The deposits are highly amorphous, low melting and surface adherent. They are composed mostly of compounds also found by extraction of the thermally stressed fuel; that is; alcohols, aldehydes, ketones, substituted furanones and carboxylic acids. The concentrations

of these products, and of the gums, increase smoothly with time.

The alternate pathway is characteristic of straight-run fuels and model compounds doped with phenols. These fuels rapidly form high melting, non-adherent, crystalline solids (with little or no gum) and oxidize slowly, seemingly with an induction period under flask test conditions. The solids have been analyzed by our sequential thermal desorbtion/pyrolysis and gas chromatography-mass spectrometry[4]. The overwhelmingly predominant functional compounds found are phenols. The same observation has also been made for solid deposits obtained from the fuel feed tube and from the spray ring of a military engine augmenter. (We feel that this represents an original success in separating and identifying molecular components of real engine deposits.) We have also found high concentrations of phenols (over 250 mg/L) in unstressed samples of these fuels.

Hazlett[5] first indicted phenols in deposit formation in 1986 but the distinctively different behavior of hydrotreated vs. straight-run fuels and the correlation of deposition to initial phenol content of fuels emphasize the importance of phenolic oligomerization as a mechanism of deposit formation. Heneghan,[6] in a matrix of thermal stability tests and analytical tests, found that phenol content was an excellent predictor of thermal stability.

Most investigators report elevated levels of nitrogen and sulfur as well as oxygen in deposits. The literature contains exotic speculations that deposits are initiated by extractable insoluble maromolecular species as well as particular compounds such as phenalenes, phenalenones, indoles, quinolines, and many others. We believe that the mechanism of solid, insoluble deposits from straight-run fuels is straight-forward and simple, at least for straight-run fuels, as compared to such speculations. Kauffman[2] has proposed a very reasonable mechanism in which a S•N compound is incorporated into bulk deposits by phenol radicals. We[1] have found that thiophenols and thiols promote deposition while sulfides and thiophenes have little effect on deposition in a model fuel containing no nitrogen. Motohashi[7] has done an extensive study of pretreatment and deposits from a light-cycle-oil. He found that caustic extraction was second only to hydrotreating for reduction of deposits. Acid extraction had little or no effect but methanol extraction and clay did

49-4

reduce deposition somewhat. Deposits contained elevated levels of N and S as well as O. The surprising finding was that filtrate of aged fuel did not show a large reduction in phenol content.

We speculate that phenols are formed from alkyl substituted aromatics with an alpha hydrogen. Air oxidation then forms alpha-arylhydroperoxides which will decompose to phenols thermally or by acid catalysis. Since the major commercial route to phenol is air oxidation of cumene followed by acid decomposition of cumene hydroperoxide to acetone and phenol.

Most straight-run fuels will contain thiols and thiophenols that can be oxidized to sulfenic or sulfonic acids in sufficient concentration to catalyze arylhydroperoxide decomposition to phenols. The phenols may have to reach a threshold concentration before oxidative phenolic coupling can occur and for a relatively stable phenol concentration to exist as fuels are thermally stressed and deposits increase.

In this work we show that a thermally stable, hydrotreated, fuel will produce the same deposition behavior as a thermally unstable straight-run fuel when doped with thiophenol. We have also tried to more specifically elucidate the exact nature of the mechanism by the use of simple model system and by thermal desorbtion/pyrolysis-GC-MS analysis of solid deposits from real engines and from a flight line heater. Much of this later work is incomplete and/or inconclusive due to a series of equipment failures of both the thermal stress apparatus and the analytical instrumentation.

Experimental

Fuels used were highly characterized samples from WL/POSF stocks. All solvents and chemicals were 99+% grade from Aldrich Co. and used as received. Oxidative thermal stress was achieved with a commercial isothermal corrosion oxidative test (ICOT) apparatus at temperatures of 140-180°C, 5 or 8 hr. times, all at 100 mL/min. flow of dried and filtered air. Solids were collected on Osmonics[R] 0.45 micron 47 mm filters under vacuum and acetone washed, dried 20-30 torr, 75°C for 24 hours. 5.0 mL portions of filtrate were extracted with 1.0 g J & W brand silica gel solid phase extraction cartridges. Cartridges were conditioned with 2x5 mL portions methanol, methanol/ acetone, acetone/heptane and heptane. Extract of filtrate was washed with 3x3 mL heptane, cartridge lightly air dried and eluted with 3.0 mL

methanol. Eluant was evaporated with dry nitrogen, weighed and immediately rediluted to 1.0 mL with methanol.

Analysis: Hewlett Packard 5890 series II GC/5890 MS with A autosampler. MS scanned 35-550 m/z, six minute solvent delay. 1.0μ splitless injection, 280 °C injection port and transfer line. 50 meter x 0.25 mm x 0.5 μm J & W DB-5 MS column at 30 psi head pressure. Purge time 0.5 min., 2 min. at 60 °C, 2 °C/min. to 250 °C, 10°c/min. to 280 and 10 min. final hold. Solid samples by thermal desorbtion/pyrolysis with CDS model 1000 Pyroprobe[R]. Coil probe, 2 x 16 mm quartz tube. Approximately 10 mg of dry sample for sequential thermal desorbtion and pyrolysis at 200, 280, 450, 750 and 1100 °C. Interface at 200 and 280 °C for first two samples, probe fired immediately upon attaining interface temperature. Interface at 325-330 °C for final three pyrolysis runs. Probe fired 99.9 seconds in all cases and GC run started immediately after probe firing. MS scanned 15-550 m/z. GC purge time four minutes. Initial temperature @ -50 °C for six minutes, 10 °C/min. to 50 °C, then three °C/min. to 280 °C and 17.33 min. hold for 110 min. Total run time with the same column as soluble sample runs.

Results and Discussion

A. Effect of Different Sulfur Compounds on Deposition

Table one gives the results of different representative sulfur compounds on deposit formation by a simple, surrogate (for JP-8) fuel that has also been doped with a phenol[1]. In this table, "gums" are acetone soluble solids that can be taken to represent oxidation products and "solids" are acetone insoluble substances that more closely represent deposits found in real engines. Of the seven sulfur compounds used, the 2-ethylthiophenol and 3,4-dimethylthiophenol dramatically increased the amount of solids formed and decrease the formation of gums. We interpret this to mean that these compounds act as antioxidants but also as deposition enhancers. We believe that this experiment needs to be completed by altering the composition of the surrogate to include some distinctive substituted aromatic compounds and by deleting the phenol depant. Such experiments would deal primarily with the effect of the sulfur compounds on phenol formation and could result in kinetic studies prior to dimerization of phenols.

B. Analysis of Real Deposit Samples

Deposits were collected from a flight line heater, the inner liner of a J-69 engine and the nozzle of a J-85 engine. Attempts have been made to characterize these deposits by the thermal desorbtion/pyrolysis-GC-MS technique. The attempts at analysis have been repeated for all three samples but little information has been gained. The samples all seem very "mature" thermally and composed mostly of inorganic carbon. Weight loss in the stepwise thermal desorption and pyrolysis of these samples is about 5% for the J-69 sample, 20% for the J-85 sample and 28% for the flight line heater sample. Both the J-69 and J-85 samples gave chromatograms characteristic of adsorbed fuel when thermally desorbed at 200 °C and essentially nothing at higher temperature thermal desorbtion and pyrolysis. The heater sample had apparently been acetone washed. The 200 ° Thermal desorbtion chromatogram shows a large amount of acetone and little else. Thermal desorbtion at 280 °C and pyrolysis at 450 °C gave traces only of low levels of benzene, tolutene and xylenes. Pyrolysis at 750 °C did produce a chromatogram with peaks identifiable as phenol, cresoles and two carbon phenols as well as much larger quantities of the unsubstituted aromatics. The chromatogram was similar to those produced by deposits from a more advanced engine[4] that first suggested phenolic coupling as the precursor to solid deposits. The presence of phenols and aromatic compounds only in these chromatograms strongly suggest that the deposits were initiated by phenolic coupling, analogous to some commercial polymers, followed by condensation of the oligomers to unsubstituted aromatic networks.

C. Doping experiments

A great deal of work, both at WL/POSF and by other contractors, has been done with two POSF fuels; POSF 2747, a hydrotreated fuel that is nearly as thermally stable as JP-TS, and POSF 2827 - a straight-run fuel that forms heavy deposits when thermally stressed.

In our previous work with these fuels[4] POSF 2747 has proved to oxidize very quickly but it did not produce solid, acetone insoluble deposits. The oxidized products from soluble gums, produced by POSF 2747, consisted predominantly of alcohols, ketones and alkenes. The polar products extracted with solid phase cartridges consisted of alcohols, ketones, some carboxylic acids and essentially a homologous

49-7

series of 5-alkyl-2,3-dihydrofuranones with C-1 to C-8 or higher alkyl substituents.

In contract, extracts of stressed POSF 2827 contained almost exclusively phenols. Thermally desorbed deposits formed by POSF 2827 were all aromatic with essentially the only functional group being phenol.

Figure one shows the chromatograms obtained from step wise (sequential) thermal desorbtion and pyrolysis of POSF 2747 doped with 300 ppm thiophenol and stressed 5 hours at 180 °C with an airflow of 100 mL/minute. These deposits were hard, crystalline and acetone insoluble. Deposits from undoped POSF 2747 are amorphous and acetone soluble. Table two is the peak identification for the chromatograms in Figure one. Aside from the thiophenol dopant and oxidation products of the thiophenol, the only organic functional group present is phenol and the composition of the deposits is very similar to that of deposits stressed POSF 2827. The fact that doping with a catalytic amount of thiophenol completely changes the physical nature and the chemical composition of deposits from stressed POSF 2747 is a very strong argument for the hydroperoxide-phenol-phenolic coupling mechanism for deposit formation.

The thiophenol, diphenylsulfide and diphenyldisulfide found in the chromatograms do not seem to account for all of the thiophenol added. Sulfonic and/or sulfuenic acids are probably present in the deposits but not detectable by gas chromatography. Authentic compounds should be obtained for development of an analytical method for them. If formation of acid sulfur compounds could be followed with time, in a fuel matrix, it would very strongly support the phenolic mechanism for deposit formation.

The mechanism of deposit formation from fuels is very important. In spite of much recent evidence to the contrary , there is still a strong tendency of concerned personnel to equate "oxidizing ability" or rate and extent of oxidation of a fuel to the propensity of that fuel to form solid deposits in use. The firm establishment of the phenolic mechanism would allow the deposition tendency of a fuel to be determined by several rapid, precise and inexpensive analytical methods.

Acknowledgements

References

1) Schulz, W.D. and Gillman, A.P., "Effect of Sulfur Compounds on a Surrogate JP-8 Fuel": Preprints, ACS Div. of Fuel Chemistry 39 (3), 943 (1994).

2) Kauffman, R.E., "The Effects of Different Sulfur Compounds on Jet Fuel Oxidation and Deposition": ASME Preprint 95-GT-222 (1995) (accepted for publication, Transactions of the ASME).

3) Heneghan, S.P., Locklear, S.L., Geiger, D.E., Anderson, S.D., and Schulz, W.D., "Static Tests of Jet Fuel Thermal and Oxidation Stability", AIAA J. Prop. Power 9, 5 (1993).

4) Schulz, W.D., "The Role of Phenols in Turbine Engine Deposit Formation": Preprints, ACS Div. of Petroleum Chemistry 39, (1), 30 (1994).

5) Hazlett, R.N. and Power, A.J., "The Role of Phenols in Distillate Fuel Stability": Proceeding 2nd International Conference on Long-Term Storage Stabilities of Liquid Fuels", San Antonio, Texas, July 29-Aug 1, 1986, pp 556-569.

6) Heneghan, S.P. and Kauffmann, R.E., "Analytic Tests and their Relation to Jet Fuel Thermal Stability", Proceedings Fifth Annual Conference on Stability and Handling of Liquid Fuels, Rotterdam, the Netherlands, Oct 3-7, 1994.

7) Motohashi, K., Nakazono, K. and Oki, Masami: "Storage Stability of Light Cycle Oil: Studies for the Root Substance of Insoluble Sediment Formation:, 5th International Conference of Stability and Handling of Liquid Fuels", Rotterdam the Netherlands, Oct. 3-7, 1994.

Table 1. Gravimetric Analysis of Surrogate JP-8 (JP-8S) doped with Sulfur Compounds.

Stress Sample	Gums	Solids	Total	SiOH Extract*
Surrogate + Dopant		Mass % of Fuel		mg/5mL Fuel
1 % 2-propylphenol/0.5 % dibenzothiophene	4.70	0.46	5.00	110.20
1 % 2-propylphenol/0.5 % 2-ethylthiophene	4.13	0.58	4.71	210.20
JP-8S (No dopant)	3.32	0.06	3.38	122.70
1 % 2-propylphenol/0.5 % phenylsulfide	2.43	0.29	2.72	191.50
1 % 2-propylphenol	1.71	0.16	1.87	155.80
1 % 2-propylphenol/0.5 % 2-ethylthiophenol	0.57	0.91	1.48	29.00
1% 2-propylphenol/0.5 % 3,4-dimethylthiophenol	0.38	1.30	1.68	32.00
1 % 2-propylphenol/0.5 % phenylethylmercaptan	0.35	0.46	0.81	34.10
1 % 2-propylphenol/0.5 % hexanethiol	0.29	0.43	0.72	21.60

*SiOH extract: Represents values for "soluble gums".

Figure 1. Total Ion Chromatograms of Solid Deposit formed from POSF 2747 Jet A-1 Fuel stressed 5 hrs. at 180 °C, with 100 mL/min. air sparge. Sequential thermal desorbtion and pyrolysis at 200, 300, 500 and 800 °C of the same sample.
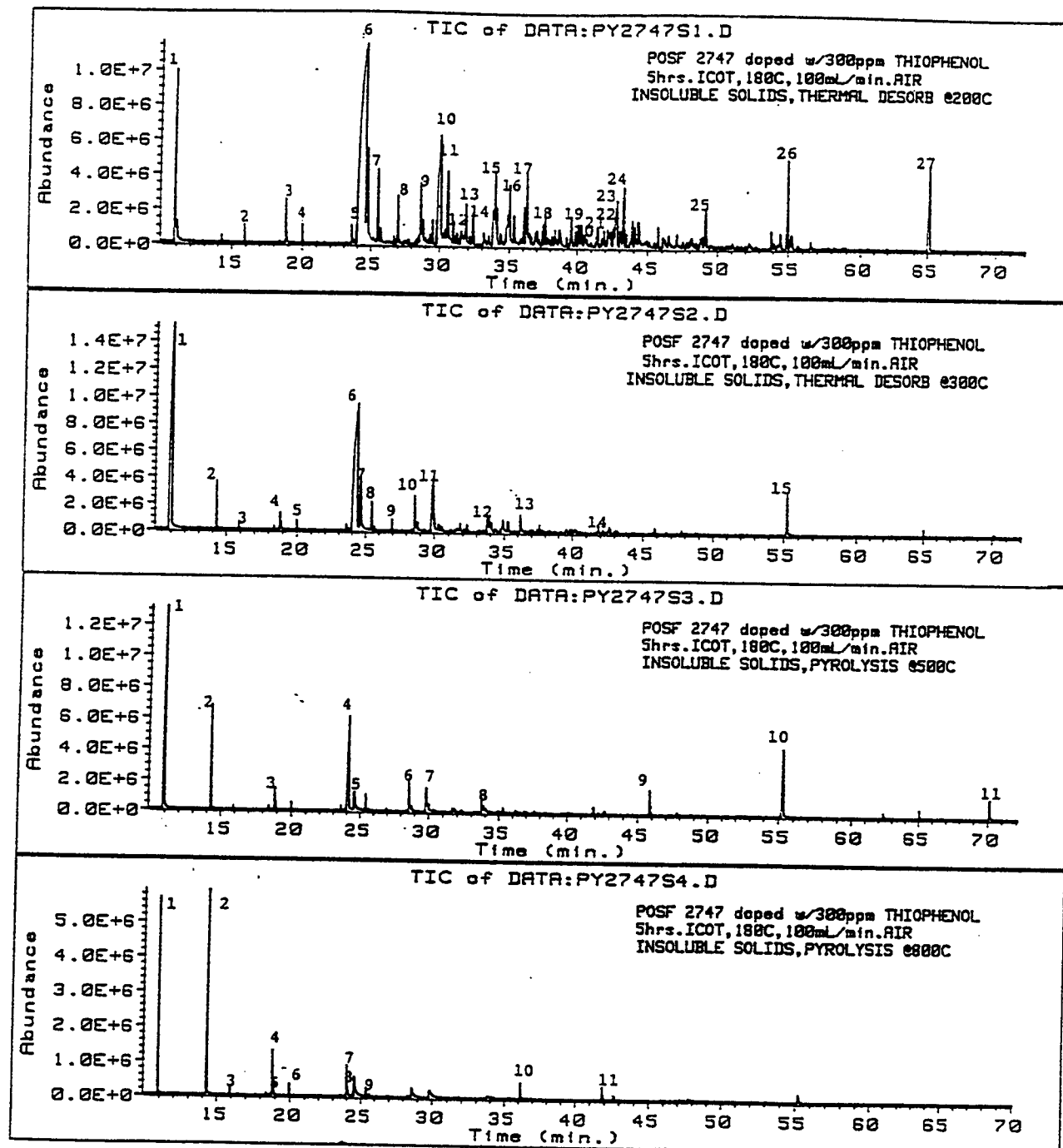
Table 2. Compound Identification for Thermal Desorbtion. Pyrolysis of Solid Deposits from POSF 2747 Doped With 300 ppm Thiophenol and Stressed 5.0 hrs. @ 180 °C.

A: PY2747S1.D. Compound from Thermal Desorbtion @ 200 °C.

Peak #	Ret. Time (min.)	Compound
1	10.92	Benzene
2	15.91	Hexamethyltrisiloxane
3	18.82	Xylene
4	19.98	Xylene
5	23.59	1-ethyl-2-methylbenzene
6	24.59	Benzenethiol
7	25.84	Trimethylbenzene
8	27.01	Trimethylbenzene
9	28.67	Cresol
10	30.06	Cresol
11	30.60	Methylthiobenzene
12	31.87	2-methylbenzofuran
13	32.37	Tetramethylbenzene
14	33.96	Dimethylphenol
15	35.04	Dimethylphenol
16	36.29	Ethylphenol
17	37.62	Dimethylbenzofuran
18	38.37	Trimethylphenol
19	39.55	3-ethyl-5-methylphenol
20	40.09	Trimethylphenol
21	41.48	Substituted Phenol
22	42.83	Mixed
23	43.31	Mixed
24	49.14	Substutited Phenol
25	53.87	Dimethylethylbenzene
26	54.99	Hexadecane
27	65.17	Diphenyldisulfide

B: PY2747S2.D. Thermal Desorbtion @ 300 °C.

Peak #	Ret. Time (min.)	Compound
1	11.04	Benzene
2	14.28	Methylbenzene
3	15.91	Cyclotrisiloxane
4	18.81	Xylene
5	19.97	Xylene
6	24.37	Benzenethiol
7	24.62	Phenol
8	25.45	Trimethylbenzene
9	26.93	Trimethylbenzene
10	28.58	Cresol
11	29.90	Cresol
12	33.75	Dimethylphenol
13	36.17	Dimethylphenol
14	41.85	Methylnaphthalene
15	55.25	1.1'-thiobisbenzene

C: PY2747S3.D. Pyrolysis @ 500 °C.

Peak #	Ret. Time (min.)	Compound
1	10.97	Benzene
2	14.31	Methylbenzene
3	18.82	Dimethylbenzene
4	24.14	Benzenethiol
5	24.58	Phenol
6	28.58	Cresol
7	29.82	Cresol
8	33.74	Dimethylphenol
9	45.85	1,1'-biphenyl
10	55.29	1.1'-thiobisbenzene
11	70.13	Thianthrene

D: PY2747S4.D. Pyrolysis @ 800 °C.

Peak #	Ret. Time (min.)	Compound
1	10.88	Benzene
2	14.28	Methylbenzene
3	15.89	Cyclotrisiloxane
4	18.71	Xylene
5	18.79	Dimethylbenzene
6	19.95	Ethylbenzene
7	23.96	Benzenethiol
8	24.52	Phenol
9	25.37	1-ethyl-4-methylbenzene
10	36.13	Naphthalene
11	41.83	Methylnaphthalene

MODELING, ANALYSIS, AND DESIGN OF BLADED DISKS FOR ALLEVIATION OF HIGH CYCLE FATIGUE IN GAS TURBINE ENGINES

Joseph C. Slater
Assistant Professor
Department of Mechanical and Materials Engineering

Wright State University
3640 Colonel Glenn Highway
Dayton, OH 45435
jslater@valhalla.cs.wright.edu

MODELING, ANALYSIS, AND DESIGN OF BLADED DISKS FOR ALLEVIATION OF HIGH CYCLE FATIGUE IN GAS TURBINE ENGINES

Joseph C. Slater
Assistant Professor
Department of Mechanical and Materials Engineering

Abstract

This report discusses methods for modeling and analyzing bladed disk assemblies common to gas turbine aircraft. An introduction to symmetric (tuned) bladed disk dynamics is presented. Special attention is paid to a literature review of studies of modeling of bladed disks, modeling of blade mistuning, and the effect of mistuning on mode localization and flutter, with a stronger emphasis being placed on forced excitation. A summary of results is given, and suggestions for future work are made.

MODELING, ANALYSIS, AND DESIGN OF BLADED DISKS FOR ALLEVIATION OF HIGH CYCLE FATIGUE IN GAS TURBINE ENGINES

Joseph C. Slater

Introduction

Gas turbine engines continue to be plagued by failures due to poorly understood vibration phenomenon. When a bladed disk such as those found in the fan, compressor, and turbine sections of a gas turbine engine, is excited by a harmonic force, high amplitude resonance may occur as a result of the lack of significant damping mechanisms. If the bladed disk is perfectly symmetrical, it will have a large number of repeated natural frequencies. The mode shapes corresponding to these frequencies are periodic around the disk. When the bladed disk is spinning, a more appropriate way of thinking about the motion is to consider the two repeated modes as traveling waves[6]. This way, the motion of the disk may be considered to be the combination of a forward travelling wave and a backward traveling wave, each travelling with some period T. A forward traveling wave is a wave traveling in the same direction as the disk is rotating, while a backward traveling wave is a wave traveling opposite the direction of rotation of the disk. If the blades are excited with a period corresponding to that of the wave speed, then the wave amplitude becomes very large, corresponding to resonance in the non-rotating system.

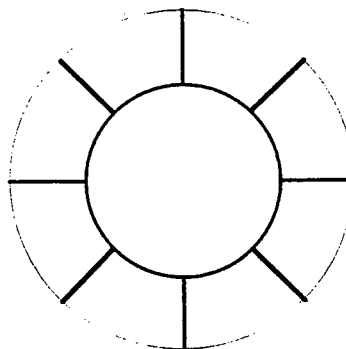Take for instance the following bladed disk where the radial lines represent blades of the disk.

Figure 1: Undeformed Bladed Disk

For the sake of illustration, shortening or lengthening of the radial line denotes deflection of a blade. The faint circle in the proceeding figure represents the nominal position of the undeformed blade. In a real bladed disk, the important deflections of the blades are usually bending, twisting, and combinations of the two – motions very similar to those of a cantilevered blade. For this illustrative example the deflection is represented as axial shortening or lengthening. This represents a simplification of the blade dynamics to those of a single beam-like mode.

Since this is an eight degree of freedom model, eight linear modes can be expected. For the perfectly tuned system, these modes will be repeated and symmetric (except for two of the modes) as shown below. The gray lines represent the nodal diameters (lines) along which there is no deflection.
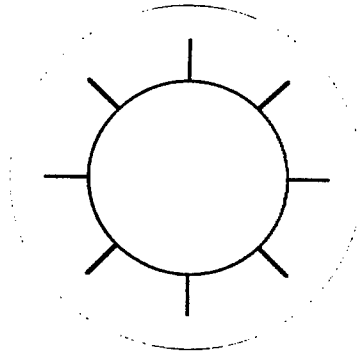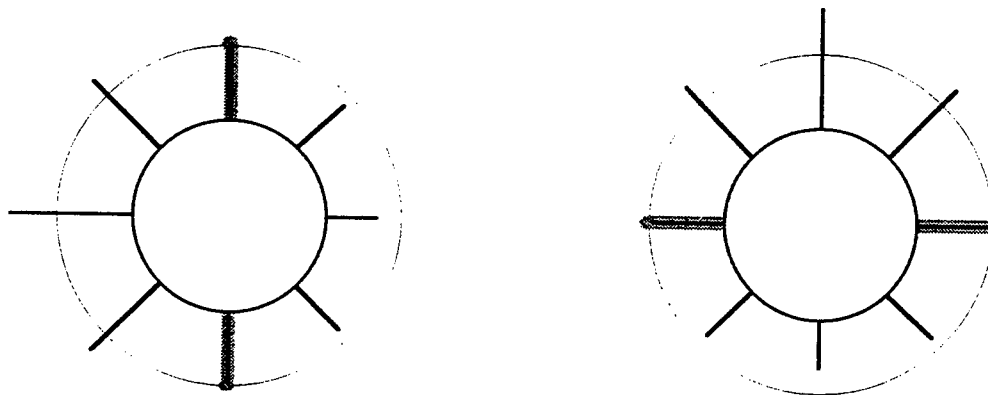
Figure 2: First Mode (Zero nodal lines)

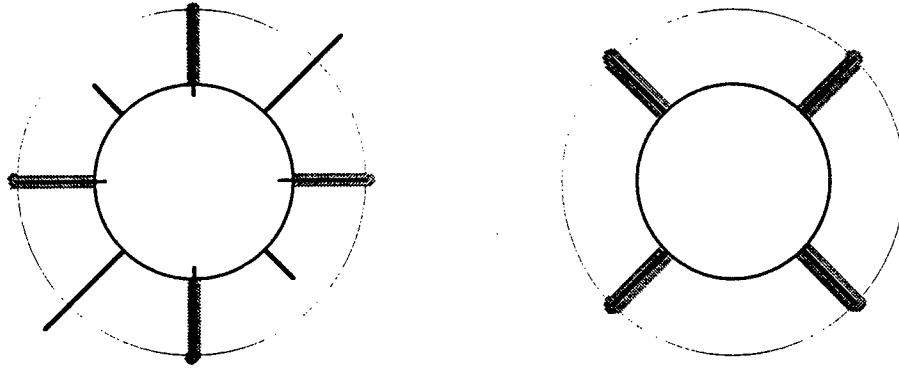**Figure 3: Second and Third Modes (Repeated, with 1 nodal diameter)**

Figure 4: Fourth and Fifth Mode (Repeated, with 2 nodal diameters)
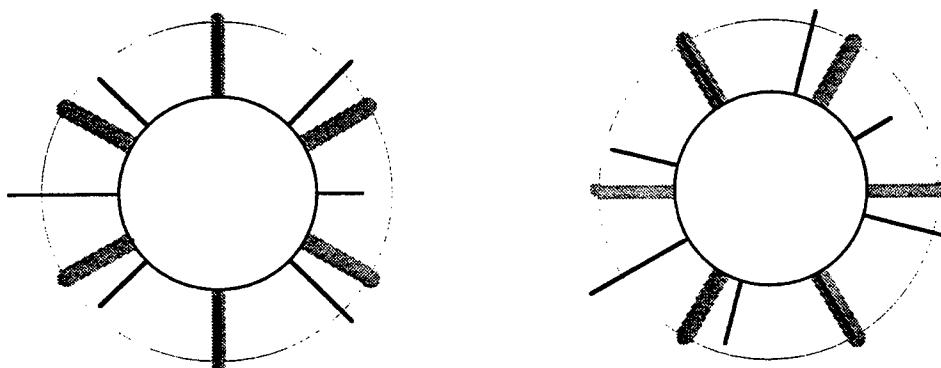

Figure 5: Sixth and Seventh Mode (Repeated, with 3 nodal diameters)
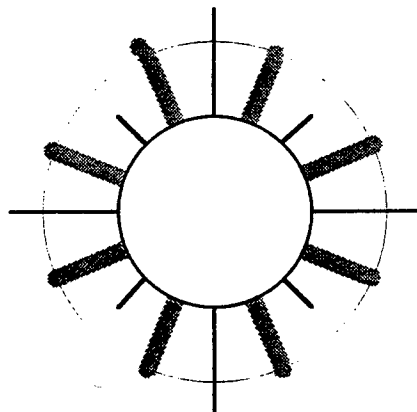

Figure 6: Eighth Mode (Not repeated, 4 nodal diameters)

It is reasonable to expect the modes to be repeated in this fashion when considering the symmetry involved in the structure. Observing modes two through seven, it should be noticed that each mode in a pair is identical to the other when rotated through some angle. When any two mode shapes in a pair are combined, they yield the same mode

shape rotated by some angle. If we take motion in both mode shapes of a repeated mode and combine them in phase, the motion appears to be simply the motion of a single mode. However, when the two modal motions are slightly out of phase, then the motion appears to be a traveling wave, traveling clockwise or counter clockwise depending on the relative phase. Since the wave may be expressed as the combination of two independent modal motions, its speed is proportional to the natural frequency of the mode.

Consider a wave caused by motion of the fourth and fifth modes. The motion is the sum of the two modal motions

$$x(t) = v_1 \sin(\omega t) + v_2 \sin(\omega t + \phi) \tag{1}$$

where $x(t)$ is the vector representing the displacements, v_1 and v_1 are vectors representing the mode shapes, ω is the natural frequency of the fourth and fifth modes, and ϕ is the phase difference between the motion of the fourth and fifth modes. Observing equation (1), $x(t)$ repeats every period of $T=2\pi/\omega$ (as well as multiples of T). Looking at Figure 4, this means that the wave must travel 90° to reach the same state in one period T. The wave speed is thus 90°/T.

Intuitively, the best way to excite such a motion is to apply two positive forces 180° away from each other and spin them about the disk at a speed equal to that of the wave speed. In a gas turbine engine, this is exactly what happens except that the force is stationary and the disk rotates. The forces are caused by the aerodynamic effects of stators and other periodic obstructions to the free flow of gases through the duct. Thus there are certain speeds of rotation of the bladed disk at which resonance occurs. This is demonstrated on the Campbell diagram in Figure 7.

The nearly horizontal lines represent the natural frequencies of the bladed disk as a function of engine speed and are labelled 0nd, 1nd, and so on meaning zero nodal diameters, 1 nodal diameter. The lines labeled 1/rev, 2/rev..., represent the frequency of the excitation force acting on the disk if it is applied more than one time per revolution. For the previous example, we were concerned with the 2 per rev. excitation because the mode shape dictates the ideal forcing distribution to be two forces per revolution. Similarly, for the third pair of repeated modes the ideal forcing distribution is three forces per revolution.
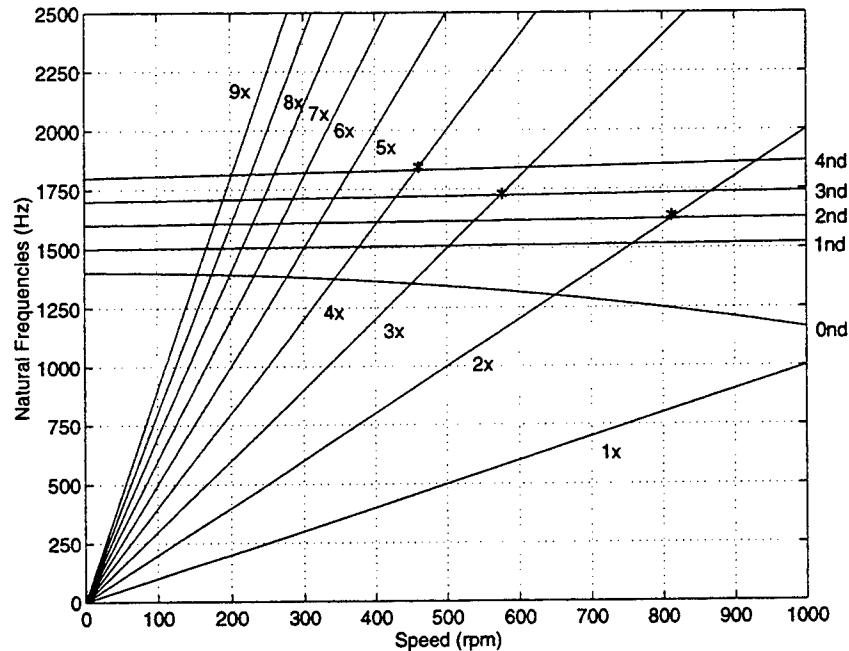
Figure 7: Typical Campbell Diagram

These results are typical of tuned analysis of bladed disks. Depending on the configuration upstream of the bladed disk, the excitations of concern may be any one of, or combination of the, *nx* excitations.

Survey of Tuned Analyses

Ewins and Imregun[16] show that pocketed bladed disk assemblies (ones with discontinuous shrouds) have modes quite different from continuously shrouded assemblies. The modes are complex in that they contain several nodal diameter components. This has significant implications with respect to forced excitations because the modal impurity makes the mode much more easy to excite at multiple rotation speeds.

In Crawley and Makadon[11], a Rayleigh-Ritz model of a bladed disk is generated in order to study blade stagger angle dependence on the coupling between the in plane and out of plane motion of bladed disks. The model is also experimentally verified. The blade model includes all six rigid body degrees of freedom of the disk, elasticity constrained by the shaft. The elastic motion of the disk is modeled by a truncated series of zero model circle disk modes. The blade is modelled as having a single heading mode only. It is shown that the first mode is a twisting mode of the bladed disk moving almost rigidly on the shaft, while the next two modes are similar bending modes of the bladed

disk on the shaft. The rest of the modes are typical of the aforementioned modes. These modes show up to a 70% variation in their natural frequencies depending upon the blade stagger angles.

Mead[36] studied wave propagation in periodic systems for mono- and multiple-coupled systems. For a bladed disk, the blades are the "systems" while the disk is the weak coupling between them. A result of this work is that each set of n modes of a bladed disk ($n=8$ for the introductory example) is bounded on the low side by the corresponding pinned-(or, rather, "ball-jointed" to include twisting compliance) free natural frequency and on the high side by the corresponding cantilevered beam frequency.

Minas and Kodiyalam[38] derive a method for determining the natural frequencies of a tuned bladed assembly using the Finite Element Model of a single blade with its corresponding symmetric section of the hub ($360°/n$ of the hub). The modeling technique is based on the assumption that the interfaces between adjacent hub protistans have a harmonic pattern as discussed in Mead[36], Lane[33], Kaza and Kielb[28], and Wei and Pierre[63]. This motion, however, is based on the coupling of systems through a single coordinate only. Also, since the number of blades is not typically a multiple of the number of modal lines in a given mode, and the effects of stagger angle vary linearly with amplitude, it is likely that this assumption is at best approximate in the tuned case. The derivation is also equivalent to that of Swaminadham, Soni, Stange, and Reed[58].

Prohl[52] and Weaver and Prohl[62] developed a method for the determination of natural frequencies of banded steam turbine buckets without making an experimental comparison. Their work represents the first extensive computer analysis of a bladed disk assembly.

Swaminadham, Soni, Stange, and Reed[58] generated a finite element model of a complex 20 bladed disk using NASTRAN. The bladed disk was integrally machined from a solid titanium block. The blades were tapered, severely twisted, swept back, had spanwise and chordwise variable thickness, and curved camber. Several attempts were made by the authors to tune the analytical model to match the results of the experimental results to no avail. The best results contained errors of approximately 30% between analytical and experimental results. Blair[5] suggested that this error is due to an incompatibility of the element types used in the model.

Survey of Mistuned Bladed Disk Analyses

Under certain conditions, due to the high flexibility of the blades relative to the disk, bladed disks can become very susceptible to mode localization as a result of blade mistuning (the variation of dynamic properties amongst the blades attached to the disk). The development of analytical methods for predicting the natural frequencies and mode shapes of tuned bladed disks has advanced sufficiently for use as design tools (although Swaminadham, Soni, Stange, and Reed[58] demonstrate that work still needs to be done). What is lacking is the ability to predict the sensitivity of a design to mode localization that may result as a function of some specified mistuning, and knowledge of how to design bladed disks that are insensitive to blade mistuning.

Statistical Approaches to Mistuned Bladed Disk Analysis

Griffin and Hoosac[21] performed simulations of a simplified model to generate a large number of samples from which to draw statistical conclusions. The model consisted of 72 three degree of freedom systems representing the 72 blades of a bladed disk. The systems were coupled at the base by springs and each was connected to ground at the end mass by a dashpot to represent system damping. The end masses and their corresponding spring stiffnesses were varied to represent the effect of mistuning. The simulation demonstrated that under the worst case scenario, the amplitude of a single blade could easily double the normal amplitude of a similarly excited tuned system. The shape of the scatter plot seems to indicate that the worst case scenario was not achieved and that the worst case would be catastrophic. Further analysis demonstrated that the range of blade natural frequencies and not the shape of the distribution of the blade natural frequencies was the dominant influence on blade amplitudes. It was also shown that the peak blade response occurred very near to the tuned natural frequency – validating the use of tuned analysis to design for natural frequency avoidance.

Griffin[20] uses a two degree of freedom model similar to that of Basu and Griffin[3] but with more sophisticated coupling to represent aerodynamic coupling of the blades. The model was demonstrated to be capable of predicting the scatter of blade amplitudes in some but not all cases. It was also found that systems that are sensitive to mistuning also tend to be numerically sensitive to modelling inaccuracies, suggesting that a deterministic approach to analyzing mistuning effects may not be possible with existing techniques.

Murthy, Pierre, and Óttarsson[39] surmise that if the motion is considered as travelling waves, then the mode localization is a result of wave reflection at the interfaces between adjacent mistuned blades. They further suggest that an average wave transmission function ($e^{-\gamma m}$) can be used to represent the tendency of the bladed disk to exhibit localized modes. Through Monte Carlo simulation, the parameter γ can be obtained. Thus the sensitivity of a design to mistuning can be observed by changes in γ.

Deterministic Approaches to Mistuned Bladed Disk Analysis

Most of the studies in blade mistuning have been of the deterministic type, where a model is generated, mistuned, and the results analyzed. As will be shown, these results vary considerably.

Valero and Bendiksen[61] developed a three degree of freedom blade model that incorporates rotation, shroud slippage, as well as friction. The shroud friction model assumes that the entire interface slips or sticks as a unit. This approach is then formulated as a linearized eigenvalue problem. Conclusions drawn are that the shroud interface angle alters the natural frequencies and the amount of friction damping observed. Surprisingly, the effects of mistuning were independent of the shroud interface angles. The mistuning effects tended to occur in the lowest modes where less deformation occurs in the hub (and thus the coupling between the blades is less apparent). It was also noted that the highly localized modes occurred when mistuning was highly concentrated in a few blades. The authors also hypothesized that mode localization can be minimized by enhancing the interblade coupling through shrouds.

Two papers by Ewins[13,14] represent standard reference material for understanding the modeling of bladed disks. Ewins[13] shows that under some mistuning conditions, blades may suffer stress levels as high as 20% greater than in a tuned system. However, rearranging the same blades on the same hub can minimize mode localization. An optimal arrangement of blade locators is proposed, although engineers at G.E.[30] still have great difficulty in identifying the variations between a given set of blades for this purpose. Ewins[14] shows that many more resonant frequencies exist for mistuned bladed disk assemblies due to the splitting of repeated modes into independent modes. The results of Ewins[13] are in contrast to those of Dye and Henry[12] and Whitehead[64]. Dye and Henry[12] show that almost a 3-fold increase in response amplitude can occur in the presence of mistuning, while Whitehead[64] analytically shows a theoretical increase of $1 - (0.5\,(1 + \sqrt{n}/2))$.

Ewins and Han[15] use a two degree of freedom blade model to study the response of a 33-bladed disk. They show that, for this specific case, blade mistuning always results in an increase in blade amplitude and that the blade with the greatest mistune always suffers the greatest motion.

Yang and Griffin's[66] substructuring technique uses the clamped-free modes of the blades to generate a reduced model of mistuned bladed disk assemblies. They show that for a simple bladed disk assembly, the reduced model natural frequencies match the natural frequencies of the original finite element model almost perfectly, and, the peak forced response is occurs at a frequency approximately 1% higher than the "true" frequency. For the mistuned case, the full and reduced models agree well. The exception is in a case where the tuned disk exhibits frequency veering.

Irretier[26] applies a modified component mode synthesis to reduce a complete bladed disk finite element model to a smaller, more tractable problem. He shows that the manner of frequency shifting due to mistuning, and the corresponding change in mode shapes, are strongly dependent on the type of mistuning.

Kaza and Kielb[28] and Kielb and Kaza[31] used aerodynamically coupled single degree of freedom blade models for their analysis. They suggest that the effects of mistuning can be beneficial or adverse depending on the engine order of the forcing function. A significant result is that it may be possible to use designed mistuning to raise the blade flutter speed without seriously degrading the forced response, although the benefits of mistuning level off at about 5% mistune. Bendiksen[4] showed a similar result. Damping is shown to be much more effective when the blades are well tuned, which may cause problems when significant damping exists in the tuned system. A more sophisticated model showed many of the same results (Kaza and Kielb[29]).

Muszynska and Jones[40] developed a five degree of freedom blade model incorporating Coulomb shroud friction, Coulomb blade to hub friction and structural damping. Their model showed that mistuning increases the response amplitude and that appropriate design of friction dampers can reduce the response by as much as an order of magnitude as compared to a non-optimally designed friction damper. They also reported that the optimal damper design effectiveness is optimal for both the tuned and mistuned case, although the amplitude for the mistuned cases are still higher than the amplitude for the tuned case. An unexpected effect is that the friction damping, due to its nonlinear nature, causes nonlinear coupling, inducing mode localization to some degree. Vakakis[60] has shown that when nonlinear coupling exists, mode localization can occur in the absence of mistuning.

Petrov[46] combines finite element, substructuring, transfer matrix, and dynamics compliance methods to develop a complex bladed disk model including shroud, joint, material damping, aerodynamic, and cable effects (for steam turbines). Isoparametric elements are used in the joint sections to model the complex geometries. A condensation technique is applied to reduce the size of the matrices[47]. According to the author, the Fortran code is efficient enough to run on a PC. The code shows that slight mistuning drastically alters the clean transfer functions obtained for a tuned system, creating numerous resonances where only a handful previously existed.

Wei and Pierre[63] demonstrate that the sensitivity of a bladed disk to mode localization as a result of mistuning is directly related to the ratio of the mistuning strength to the coupling strength using a single degree of freedom blade model. They show that the effects of mistuning are minimal when coupling is great. When coupling is weak, however, the bladed disk is very sensitive to mistuning. Thus, a bladed disk assembly that shows a great deal of motion of the hub when moving in a mode will be less susceptible to the effects of mistuning. Since more relative motion occurs in the hub in higher modes, it seems likely that higher modes should be less susceptible to blade mistuning.

Pierre and Murthy[50] and Pierre, Smith, and Murthy[51] included aerodynamic coupling of the blades in their perturbation approach to determination of the effects of blade mistuning. Since it is shown that the low coupling between the blades is the cause of the propensity for mode localization, Pierre and Murthy apply the approach of Wei and Pierre[63] where the coupling is approached as being the perturbation of n originally independent blades with slightly different modal parameters. This is in contrast to procedures that include the mistuning as the perturbation of an originally tuned system. A heuristic explanation for why this may work is that when mode localization occurs, the blades act almost as independent structures. Since the nominal structure used by Wei and Pierre[63] is the set of independent blades, it reasonable to expect that this should yield better results. Pierre and Murthy[50] also report that blades similar in frequency tend to vibrate together in a localized mode, even when the blades between them do not show significant motion.

Óttarsson and Pierre[45] used a transfer matrix approach to model a bladed disk. The blades were modelled as one degree of freedom systems, while the coupling was through a hub represented by a concentric set of single degree of freedom systems connected to each other by springs. Mistuning was introduced as randomness in the transfer matrices. A perturbation approach was used to obtain a localization factor (a factor relating to the attenuation of travelling

waves passing from one blade to the next) for regions of high and low sensitivity to blade mistuning. The results were verified by Monte Carlo simulations. This is the same parameter used by Murthy, Pierre, and Óttarsson[39].

A summary of these results leads to the following conclusions:

1) Detailed finite element analysis of tuned bladed assemblies is prone to large errors.

2) Detailed finite element analysis of mistuned bladed assemblies is extremely costly due to the inability to apply symmetry relations.

3) The eigensolutions of even simple models are extremely susceptible to numerical difficulties as a result of the sensitivity of the eigenvectors to small parameter variations.

4) Even if detailed FEA of mistuned bladed disks yielded valid results, usefulness for design is questionable[30].

5) The most promising method of gaining a detailed finite element model that is capable of incorporating the detailed effects of blade mistuning is to apply component mode synthesis[1,7,8,9,10,18,25,32,35,43,44,57] as performed by Irretier[26].

6) Mistuning has the greatest effect when coupling between the blades is weakest. Added coupling through shrouds is likely to reduce mode localization.

7) In addition, since higher sets of bladed disk modes tend to have greater coupling between the blades, mode localization may be a phenomenon of interest only with respect to the n lowest modes of an n bladed disk.

8) Dry friction can cause mode localization to occur in the absence of blade mistuning.

9) Blade mistuning is not guaranteed to cause mode localization. The same set of mistuned blades will exhibit symmetric modes or localized modes depending on the arrangement of the blades on the hub. Little is understood about why this is the case, or what way to order the bladed to minimize this effect.

10) It is likely that blade mistuning can cause responses well above those reported in most studies. Monte Carlo simulations show distributions likely to have very sharp peaks, rather then softer peaks.

Recommendations for Future Work

In light of the observations made in this literature survey, the following are believed by the author to be the best directions for future research.

1) In light of the great degree of complexity in modeling mistuned bladed disk assemblies, and the fact that design from these finite elements models is unlikely to be a valid design tool, it is proposed that design should focus on the sensitivity of an assembly to mode localization, instead of attempting to model sets of blade mistunings incorporated into design models. This involves creating a better understanding of the relationship between blade compliance and hub compliance – how do we increase coupling between blades?

2) Studies need to be performed to verify if mode localization is an effect found only in the lowest n modes. If this is the case, the problem of design to prevent mode localization is reduced to a much smaller problem.

3) New hub designs that are not axially symmetric should be studied. Numerous papers have shown that eigenvalue veering is a very likely cause of loss of eigenstructure[2,22,23,24,42,48,59,65]. Since eigenvalue veering is a result of closely spaced modes, and axially symmetric systems have repeated modes, eigenvalue veering is likely to be unavoidable. By designing balanced hubs lacking axial symmetry it may be possible to split the repeated natural frequencies sufficiently to prevent eigenvalue veering, and thus mode localization due to blade mistuning.

<u>Suggested Reading</u>

The original published work in this area was authored by Campbell in 1924[6]. Campbell's treatise on the dynamics of bladed disk assemblies remains very relevant even today. Rao's[53] text is the only text I have been able to find detailing the modeling, analysis, and design of bladed disk assemblies. Although it is impossible to present all of the relevant work in the field of vibration of bladed disks, this text is a very good place to start understanding modern analysis techniques. Srinivasan[55] presented a survey of 46 papers on the topic in 1983, summarizing the major observations of the authors. Liessa, MacBain, and Kielb[34] reviewed a number of methods developed for modeling curved, twisted, cantilevered beams and describe the strengths and weaknesses of each. References for some other relevant work, not summarized in this writing, regarding damping[19,27], vibration absorbers[41], and measurement[56] are listed in the references.

References

1) Avitabile, P., O'Callahan, J.C., and Milani, J., "Comparison of System Characteristics Using Various Model Reduction Techniques," *Seventh International Modal Analysis Conference*, Las Vegas, Nevada, Feb. 1989.

2) Balmès, E., "High Modal Density, Curve Veering, Localization: A Different Perspective on the Structural Response," *Journal of Sound and Vibration*, Vol. 161, No. 2, 1993, pp. 358-363.

3) Basu, P., and Griffin, J.H., "The Effect of Limiting Aerodynamic and Structural Coupling in Models of Mistuned Bladed Disk Vibration," *Journal of Vibration, Acoustics, Stress, and Reliability in Design*, Vol. 108, 1986, p. 132-139.

4) Bendiksen, O.O., "Flutter of Mistuned Turbomachinery Rotors," *Journal of Engineering for Gas Turbines and Power*, Vol. 106, pp. 25-33.

5) Blair, J.J., Personal Communication, July 1995.

6) Campbell, W., "The Protection of Steam-Turbine Disk Wheels From Axial Vibration", *ASME Transactions*, Vol. 46, No. 1920, 1924, pp. 31-160.

7) Craig, Jr., R.R., *Structural Dynamics*, Wiley, New York, 1981, pp. 467-494.

8) Craig, Jr., R.R., and Chung, Y., "Generalized Substructure Coupling Procedure for Damped Systems," *AIAA Journal*, Vol. 20, March 1982, pp. 442-444.

9) Craig, Jr., R.R., "A Review of Time-Domain and Frequency-Domain Component-Mode Synthesis Methods," *Journal of Modal Analysis*, April 1987, pp. 59-72.

10) Craig, Jr., R.R., and Hale, A.L., "Block-Krylov Component Synthesis Method for Structural Model Reduction," *Journal of Guidance Control and Dynamics*, Vol. 11, Nov.-Dec. 1988, pp. 562-570.

11) Crawley, E.F., and Mokadam, D.R., "Stagger Angle Dependence of Inertial and Elastic Coupling in Bladed Disks," *Journal of Vibration Acoustics, Stress, and Reliability in Design*, Vol. 106, 1984, pp. 181-189.

12) Dye, R.C.F., and Henry, T.A., "Vibration Amplitudes of Compressor Blades Resulting from Scatter in Natural Frequencies," *Journal of Engineering for Power*, Vol. 91, 1969, pp. 182-188.

13) Ewins, D.J., "The Effect of Detuning upon the Forced Vibrations of Bladed Disks," *Journal of Sound and Vibration*, Vol. 9, No. 1, 1969, pp. 65-79.

14) Ewins, D.J., "Vibration Characteristics of Bladed Disk Assemblies," *Journal of Mechanical Engineering Science*, Vol. 15, No. 3, 1973, pp. 165-186.

15) Ewins, D.J., and Han, Z.S., "Resonant Vibration Levels of a Mistuned Bladed Disk," *Journal of Vibration, Acoustics, Stress, and Reliability in Design*, Vol. 106, April 1984, pp. 211-217.

16) Ewins, D.J., and Imregun, M., "Vibration Modes of Packeted Bladed Disks," *Journal of Vibration, Acoustics, Stress, and Reliability in Design*, Vol. 106, April 1984, pp. 175-180.

17) Fabunmi, J.A., "Forced Vibrations of a Single Stage Axial Compressor Rotor," *Journal of Engineering for Power*, Vol. 102, April 1980, pp. 322-328.

18) Gawronski, W., and Williams, T., "Model Reduction for Flexible Space Structures," *Journal of Guidance Control and Dynamics*, Vol. 14, Jan.-Feb. 1991, pp. 68-76.

19) Griffin, J.H., "Friction Damping of Resonant Stresses in Gas Turbine Engine Airfoils," *Journal of Engineering for Power*, Vol. 102, April 1980, pp. 329-333.

20) Griffin, J.H., "On Predicting the Resonant Response of Bladed Disk Assemblies," *Journal of Engineering for Gas Turbines and Power*, Vol. 110, Jan. 1988, pp. 45-50.

21) Griffin, J.H., and Hoosac, T.M., "Model Development and Statistical Investigation of Turbine Blade Mistuning," *Journal of Vibration, Acoustics, Stress, and Reliability in Design*, Vol. 106, April 1984, pp. 204-210.

22) Happawana, G.S., Bajaj, A.K., and Nwokah, O.D.I., "A Singular Perturbation Perspective on Mode Localization," *Journal of Sound and Vibration*, Vol. 147, No. 4, 1991, pp. 361-365.

23) Happawana, G.S., Bajaj, A.K., and Nwokah, O.D.I., "A Singular Perturbation Analysis of Eigenvalue Veering and Modal Sensitivity in Perturbed Linear Periodic Systems," *Journal of Sound and Vibration*, Vol. 160, No. 2, 1993, pp. 225-242.

24) Hodges, C.H., "Confinement of Vibration by Structural Irregularity", *Journal of Sound and Vibration*, Vol. 82, No. 3, 1982, pp. 411-424.

25) Hurty, W.C., "Dynamic Analysis of Structural Systems Using Component Modes," *AIAA Journal*, Vol. 3, April 1965, pp. 678-685.

26) Irretier, H., "Spectral Analysis of Mistuned Bladed Disk Assemblies by Component Mode Synthesis," *Proceedings of the Ninth Conference on Mechanical Vibration and Noise of the Design and Production Engineering Technical Conferences*, ASME, 1983, pp. 115-125.

27) Jones, D.I.G., "Vibrating Beam Dampers for Reducing Vibration in Gas Turbine Blades," *Journal for Engineering Power*, Vol. 97, 1975, pp. 111-116.

28) Kaza, K.R.V., and Kielb, R.E., "Effects of Mistuning on Bending-Torsion Flutter and Response of a Mistuned Cascade in Incompressible Flow," *AIAA Journal*, Vol. 20, No. 8, 1982, pp. 1120-1127.

29) Kaza, K.R.V., and Kielb, R.E., "Flutter of Turbofan Rotors with Mistuned Blades," *AIAA Journal*, Vol 22, No. 11, Nov. 1984, pp. 1618-1625.

30) Kielb, R.E., Personal Communications, August 1995.

31) Kielb, R.E., and Kaza, K.R.V., "Aeroelastic Characteristics of a Cascade of Mistuned Blades in Subsonic and Supersonic Flows," *Journal of Vibration, Acoustics, Stress, and Reliability in Design*, Vol. 105, Oct. 1983, pp. 425-433.

32) Kienholz, D.A., and Smith, K.E., "Admittance Modeling: Frequency Domain, Physical Coordinate Methods for Multi-Component Systems," *CSA Engineering Report*, Feb. 1994, pp. 608-614.

33) Lane, F., "System Mode Shapes in the Flutter of Compressor Blade Rows," *Journal of the Aeronautical Sciences*, Vol. 23, Jan. 1956, pp. 54-56.

34) Liessa, A.W., MacBain, J.C., and Kielb, R.E., "Vibrations of Twisted Cantilevered Plates–Summary of Previous and Current Studies," *Journal of Sound and Vibration*, Vol. 96, No. 2, 1984, pp. 159-173.

35) MacNeal, R.H., "A Hybrid Method of Component Mode Synthesis," *Computers and Structures*, Vol. 1, 1971, pp. 581-601.

36) Mead, D.J., "Wave Propagation and Natural Modes in Periodic Systems: Mono-Coupled Systems," *Journal of Sound and Vibration*, Vol. 40, No. 1, 1975, pp. 1-18.

37) Menq, C.-H., Griffin, J.H., and Bielak, J., "The Forced Response of Shrouded Fan Stages," *Journal of Vibration, Acoustics, Stress, and Reliability in Design*, Vol. 108, Jan. 1986, pp. 50-55.

38) Minas, C., and Kodiyalam, S., "Vibration Analysis of Bladed Disc Assemblies," *International Modal Analysis Conference*, 1995.

39) Murthy, D.V., Pierre, C., and Óttarsson, G., "Efficient Design Constraint Accounting for Mistuning Effects in Engine Rotors," *AIAA Journal*, Vol. 33, No. 5, 1994, pp. 960-962.

40) Muszyska, A., and Jones, D.I.G., "A Parametric Study of Dynamic Response of a Discrete Model of Turbomachinery Bladed Disc," *Transactions of the ASME*, Vol. 105, 1983, pp. 434-443.

41) Natsiavas, S., "Steady State Oscillations and Stability of Non-linear Dynamic Vibration Absorbers," *Journal of Sound and Vibration*, Vol. 156, 1992, pp. 227-245.

42) Natsiavas, S., "Mode Localization and Frequency Veering In A Non-conservative Mechanical System With Dissimilar Components," *Journal of Sound and Vibration*, Vol. 165, 1993, pp. 137-147.

43) O'Callahan, J.C., Avitabile, P.A., Riemer, R., "System Equivalent Reduction Expansion Process (SEREP)," *Seventh International Modal Analysis Conference*, Las Vegas, Nevada, Feb. 1989.

44) O'Callahan, J.C., "A Procedure for an Improved Reduced System (IRS) Model," *Seventh International Modal Analysis Conference*, Las Vegas, Nevada, Feb. 1989.

45) Óttarsson, G., and Pierre, C., "A Transfer Matrix Approach to Vibration Localization in Mistuned Blade Assemblies," *Proceedings of the International Gas Turbine and Aeroengine Congress and Exposition*, Cincinnati, Ohio, 1993, ASME 93-GT-115, pp. 1-20.

46) Petrov, E.P., "Analysis and Optimal Control of Stress Amplitudes Upon Forced Vibration of Turbomachine Impellers with Mistuning," *International Union of Theoretical and Applied Mechanics Symposium on The Active Control of Vibration*, Sept. 5, 1994, University of Bath, UK, pp. 189-196.

47) Petrov, E.P., "Large-Scale Finite Element Models of Blade-Shroud and Blade-Disk Joints and Condensation Technique for Vibration Analysis of Turbomachine Impellers," *Proceedings of the 7th World Congress on Finite Element Methods: "FEM: Today and the Future,"* Monte-Carlo, 1993, pp. 507-513.

48) Pierre, C., "Mode Localization and Eigenvalue Loci Veering Phenomena in Disordered Structures," *Journal of Sound and Vibration*, Vol. 126, 1988, pp. 485-502.

49) Pierre, C., and Dowell, E.H., "Localization of Vibrations by Structural Irregularity," *Journal of Sound and Vibration*, Vol. 114, 1987, pp. 549-564.

50) Pierre, C., and Murthy, D.V., "Aeroelastic Modal Characteristics of Mistuned Blade Assemblies: Mode Localization and Loss of Eigenstructure," *AIAA Journal*, Vol. 30, No. 10, 1992, pp. 2483-2496.

51) Pierre, C., Smith, T.E., and Murthy, D.V., "Localization of Aeroelastic Modes in Mistuned High-Energy Turbines," *Journal of Propulsion and Power*, Vol. 10, 1994, pp. 318-328.

52) Prohl, M.A., "A Method for Calculating Vibration Frequency and Stress of a Banded Group of Turbine Buckets," *Transactions of the ASME*, Jan. 1958, pp. 169-180.

53) Rao, J. S., *Turbomachine Blade Vibration*, Wiley, New York, 1991.

54) Rzadkowski, R., "The General Model of Free Vibrations of Mistuned Bladed Discs, Part I: Theory," *Journal of Sound and Vibration*, Vol. 173, 1994, pp. 377-393.

55) Srinivasan, A.V., "Vibrations of Bladed Disk Assemblies- A Selected Survey," *Journal of Vibration, Acoustics, Stress, and Reliability in Design*, Vol. 106, 1984, pp. 165-168.

56) Srinivasan, A.V., and Cutts, D.G., "Measurement of Relative Vibratory Motion at the Shroud Interfaces of a Fan," *Journal of Vibration, Acoustics, Stress, and Reliability in Design*, Vol. 106, 1984, p. 106.

57) Su, T., and Craig, R.R., "Krylov Model Reduction Algorithm for Undamped Structural Dynamics," *Journal of Guidance Control and Dynamics*, Vol. 14, Nov.-Dec. 1991, pp. 1311-1313.

58) Swaminadham, M., Soni, M.L., Stange, W.A., and Reed, J.D., "On Model Generation and Modal Analysis of Flexible Bladed-Disc Assemblies," *Bladed Disk Assemblies*, ASME Vibration Conference, Cambridge, MA, Sept. 27-30, 1987, pp. 49.

59) Triantafyllou, M.S., and Triantafyllou, G.S., "Frequency Coalescence and Mode Localization Phenomena: A Geometric Theory," *Journal of Sound and Vibration*, Vol. 150, 1991, pp. 485-500.

60) Vakakis, A.F., "Non-Similar Normal Oscillations in a Strongly Non-Linear Discrete System," *Journal of Sound and Vibration*, Vol. 158, 1992, pp. 341-361.

61) Valero, N.A., and Bendiksen, O.O., "Vibration Characteristics of Mistuned Shrouded Blade Assemblies," *Journal of Engineering for Gas Turbines and Power*, Vol. 108, 1986, pp. 293-299.

62) Weaver, F.L., and Prohl, M.A., "High-Frequency Vibration of Steam-Turbine Buckets," *Transactions of the ASME*, Jan. 1958, pp. 181-194.

63) Wei., S.T., and Pierre, C., "Localization Phenomena in Mistuned Assemblies with Cyclic Symmetry Part 1: Free Vibrations," *Journal of Vibration, Acoustics, Stress, and Reliability in Design*, Vol. 110, 1988, pp. 429-438.

64) Whitehead, D.S., "Effects of Mistuning on the Forced Vibration of Blades with Mechanical Coupling," *Journal of Mechanical Engineering Science*, Vol. 18, No. 6, 1976.

65) Wu, G., "Letter to the Editor: Free Vibration Modes of Cyclic Assemblies With A Single Disordered Component," *Journal of Sound and Vibration*, Vol. 165, 1993, pp. 563-570.

66) Yang, M.-T., and Griffin, J.H., "A Reduced Order Approach for The Vibration of Mistuned Bladed Disk Assemblies," *International Gas Turbine and Aeroengine Congress & Exposition*, 1995, ASME Paper No. 95-GT-454.

A STUDY OF PULSED LASER DEPOSITION
OF SILICON CARBIDE THIN FILMS

Ronald Birkhahn
Graduate Student Associate
Department of Materials Science

Andrew J. Steckl
Professor
Department of Electrical and Computer Engineering

University of Cincinnati
899 Rhodes Hall
P.O. Box 210030
Cincinnati, OH 45221-0030

Final Report for:
Summer Faculty Research Program
Wright Laboratory

September 19, 1995

A STUDY OF PULSED LASER DEPOSITION OF SILICON CARBIDE THIN FILMS

Ronald Birkhahn
Graduate Student Associate
Department of Materials Science

Andrew J. Steckl
Professor
Department of Electrical and Computer Engineering

ABSTRACT

We investigated the validity of using pulsed laser deposition (PLD) to deposit single crystal epitaxial SiC films during the Summer Research Program at Wright Patterson Air Force Base. After modification of the substrate stage for heating capability, SiC was pulsed-laser-deposited onto (001) Si, SiC on (111) Si, and SiC substrates. X-ray diffraction (XRD) showed that the crystallinity of the deposited SiC film improved with increasing substrate temperature and reduced laser repetition rate, but remained polycrystalline overall. However, discrepancies between temperature characterization techniques prevented us from accurately determining the substrate temperature.

TABLE OF CONTENTS

FIGURES AND TABLES

Figure 1: Schematic of PLD deposition system

Figure 2: Photo of high temperature stage with substrate

Figure 3: Stage temperature vs. input power

Figure 4: Organization of experiments

Figure 5: XRD of carbonized and low pressure CVD SiC grown on Si(111) substrate before and after SiC PLD

Figure 6: Combined XRD data for PLD runs #5-7 for comparison

Figure 7: XRD of PLD-SiC grown on SiC substrates

Figure 8: Large angular scan of PLD on Lely-grown 15R-SiC substrate

Table 1: Summary of SiC PLD growth conditions

A STUDY OF PULSED LASER DEPOSITION
OF SILICON CARBIDE THIN FILMS

by
Ronald Birkhahn
Andrew J. Steckl

1. INTRODUCTION

Much work has been done with pulsed laser deposition (PLD) of SiC aiming for the ultimate goal of obtaining clean, epitaxial growth on substrates at low temperature. A list of references to the relevant literature is included with this report. To date, this has not been accomplished but several papers [Fenner(8), Rimai(17), Rimai(20)] have hinted at epitaxial growth. This "epitaxial" growth is primarily polycrystalline, sometimes including amorphous areas, with grain sizes on the order of tens of nanometers and exhibiting an oriented relationship with the substrate. The optimum growth temperatures associated with obtaining these polycrystalline "epitaxial" samples has been in the range of 800-1200°C. Most films have been grown on (001) or (111) Si substrates with one notable exception using 6H-SiC [Stan(23)]. Relatively few results have been published regarding the effect of the types of targets and laser energy. Many variables can affect the PLD growth of SiC: (a) laser parameters: energy, wavelength, pulse duration, repetition rate and deposition time; (b) substrate: type (Si, SiC), cleaning procedure, deposition temperature, and annealing. In this brief (2 month) project, we have attempted to explore a few of these parameters in our SiC PLD growth runs.

2. EXPERIMENTAL PROCEDURE

§2.1 PLD System Description

The PLD system is a high vacuum chamber designed for base pressures in the 10^{-7}-10^{-8} range. A backed turbo pump provides pressures to 10^{-6} at which time an ion getter pump takes over. The substrate and target are mounted along the long axis as seen in Fig. 1. Side ports are used for optical viewing, laser entry, sample exchange, and temperature measurement. The target is capable of rotation and the substrate stage contains the electrical feedthroughs for the molybdenum block heater capable of reaching a temperature of 1600°C. Fig. 2 shows a photograph of the high temperature stage with a Si substrate in place. We installed this high temperature stage to increase the capabilities of this PLD system, which was previously only able to run at room temperature. Fabrication of a new support cantilever was required as well as modification of the block heater to mount the small substrate samples. After installation, we

calibrated the stage temperature vs. input power from a high voltage power supply using an IRCON 2-color pyrometer (Fig.3). However, we experienced persistent problems in obtaining reliable temperature data for substrates mounted on the molybdenum block. Initial tests with the pyrometer and a chrome-alumel thermocouple attached to a piece of silicon on the heater yielded as much as a 300°C discrepancy. Subsequent tests with the thermocouple attached to the Mo block showed ambiguous results after rotation of the substrate stage. During some PLD runs, the thermocouple was clipped to the substrate and yielded a consistent 300°C difference compared to the pyrometer. This was probably due to three causes: 1) the substrate was not in full contact with the molybdenum block and presented a thermal transfer problem; 2) the pyrometer was limited by its field of focus and thus measured some of the glowing heater block; and 3) the thermocouple averaged temperature readings from the face of the substrate and the clip which acted as a heat sink. The temperature control itself, however, was consistent and reproducible.

§2.2 Sample Preparation and System Operation

Approximately 1x1 cm^2 samples used for deposition were first cleaned using standard RCA procedure for semiconductors using a base clean with NH_3OH and H_2O_2 and an acid dip of HCL and H_2O_2 at 70°C. The samples had a final HF dip before being clipped to the molybdenum heater and inserted in the chamber. The chamber was then pumped until a base pressure of low 10^{-7} or high 10^{-8} Torr is reached. Next, the sample is quickly ramped to deposition temperature and allowed to thermally equilibrate before deposition begins. A 248 nm Kr-Fl eximer laser provides the power to ablate the SiC target and to produce deposition onto the facing substrate. The laser pulses are introduced through a focusing and scanning mirror onto the target to simultaneously concentrate the power and provide a uniform deposition. The laser power was kept roughly constant for all experiments. The pulse has a 24 ns duration while the frequency is an adjustable parameter. The pulsed beam ablates the SiC target and the plume is ejected normal from surface, uniformly covering the substrate. The substrate temperature is monitored by the two-color IR pyrometer discussed above. The sample is allowed to cool after deposition and removed.

§2.3 Analysis and Characterization Equipment

After removal from the deposition chamber, the PLD layer thickness and morphology was determined by a Sloan Dektak profilometer and an optical microscope. The samples were also examined by a Philips x-ray diffractometer to determine the crystallinity of the layer and compared to previous θ-2θ rocking curves of the undeposited substrate.

3. EXPERIMENTAL RESULTS

§3.1 Experimental Conditions

Initially, we attempted PLD growth on (100) Si substrates. A second set of experiments utilized (111) Si substrates which had a single crystal 3C-SiC film grown on top. The (111) Si substrates were first carbonized at 1200°C and atmospheric pressure and then had 0.4 µm SiC film grown at 900°C and low pressure. These samples were grown in a CVD reactor at the University of Cincinnati. The other two types of samples used were 6H-SiC from Westinghouse (J23-#2) and Lely-grown 15R-SiC from Kiev, Ukraine. The organization of experiments is shown in Fig. 4.

As previously discussed in the introduction, there are many variables that could affect SiC PLD growth. The primary factors we investigated were substrate type, laser repetition rate, and temperature. During all runs the laser power was kept constant at around 230 mJ and the frequency of the laser was varied between 1 and 20 Hz. The duration of deposition was varied only to allow time for measurable SiC film to deposit. The temperature of the substrate during deposition was varied between 850 and 1200°C.

§3.2 Results and Discussion

The overall growth rate of the PLD samples varied according to the frequency of the laser. As shown in Table 1, the SiC growth rates per pulse were on the order of tenths of angstroms for the laser power utilized. This is extremely fast compared to CVD or MBE, corresponding to approximately 1mm per second of laser ablation time. A PLD study by Rimai et. al [17] reported growth rates on Si substrates using 300mJ per pulse and 800-1200°C temperatures to be still an order of magnitude larger at 2Å per pulse. The summation of all the data from the runs is listed in Table 1. The overall surface morphology determined by Dektak profilometry was fairly uniform with isolated peaks. Further examination of the PLD surface under an optical microscope revealed pits but an otherwise smooth surface layer.

The next step of analyzing the PLD films utilized the Philips x-ray diffractometer. Scans from initial PLD runs #1-3 on Si substrates had no x-ray peaks except for the substrate, indicating that the samples were amorphous in character. Figure 5 shows the results from the PLD runs on the 3C-SiC CVD growth on (111) Si. For comparison, the XRD scan in Fig. 5a was taken before PLD. All the graphs are drawn on the same scale and indicate the shifting and attenuation of the peak height with respect to the original. Run #5 (Fig. 5b) used a repetition rate of 10 Hz and substrate temperature around 850°C. Run #6 (Fig. 5c) decreased the rep rate to 5 Hz but kept the temperature of the substrate the same. The x-ray peak considerably flattened showing an amorphous character. This is a somewhat surprising result because it was expected that as you decrease the rep rate, the arriving Si and C atoms have sufficient time to rearrange on the surface to

form an oriented polycrystalline if not epitaxial film. This may have been an anomalous result considering that run #7 (Fig. 5d) had the same rep rate as #6 but a higher temperature and formed a considerably better film. A compilation of the x-ray scans of runs #5-7 are shown in Fig. 7. Note that these films are probably polycrystalline with an oriented relationship with the substrate. Another analytical technique such as transmission electron microscopy (TEM) would be necessary to confirm this.

Figure 7 displays for comparison all the PLD runs on SiC substrates along with x-ray diffraction on corresponding undeposited substrates. Run #9 exhibits a peak that is broader than the other scans for that type of substrate but the numerous amount of other peaks in the original and the scans after PLD do not allow us to draw many conclusions about the character of the films. It would appear that run #9 contained more polycrystallites. The results from the Lely-grown 15R-SiC were even more ambiguous. The original as well as the after growth sample had very sharp x-ray peaks (Fig.6d-e). A full x-ray scan of the sample (Fig. 7) revealed several peaks although the only conclusions we could draw were that there are peaks from a 15R-SiC substrate with possible evidence of 15R or 6H-SiC in a polycrystalline film. Since many of the peaks from the two different SiC polytypes overlap, it is necessary to use another analytical technique to determine the true character of the film. It is possible that an epitaxial single crystal film formed on the 15R substrate, but it is also possible that the PLD layer is completely amorphous. Both would present the same type of evidence by x-ray diffraction.

4. SUMMARY AND CONCLUSIONS

During our summer research program at Wright Patterson Air Force Base, we made progress investigating the validity of using pulsed layer deposition to deposit epitaxial films at reduced (or even room) temperatures. After installing the molybdenum heater stage to add the ability to heat the substrate during deposition, we proceeded with runs on different types of substrate under varying conditions and examined the results with x-ray diffraction. We were unable to make use of any other analytical techniques during the course of our brief stay to validate or reject the conclusions presented here. Results from those PLD runs are inconclusive at best and discouraging at worst. Gradually increasing the deposition temperature towards the 800-1200°C threshold for crystalline films found by others appeared to increase crystallinity. However, due to the discrepancy between our temperature measuring techniques for the substrate and block heater, we were unable to accurately and definitively determine true temperatures. The actual temperature might have been as much as 300°C lower that measured and would corroborate that it is not possible to attain epitaxial films below the threshold with these conditions. It is recommended that if research were to continue along this direction that the decisive answers be found for the inconclusive results (temperature, crystallinity) presented here before continuing.

REFERENCES

1. Alunovic, M., *et al.*, *Description of transfer and depostition during PLD of thin ceramic films.* ISIJ International, 1994. **34**(6): p. 507-15.

2. Balooch, M., *et al.*, *Deposition of SiC films by pulsed excimer laser ablation.* Applied physics letters, 1990. **57**(15): p. 1540-2.

3. Blouin, M., *et al.*, *Atomic force microscopy study of the microroughness of SiC thin films.* Thin solid films, 1994. **249**: p. 38-43.

4. Boily, S., *et al.*, *SiC membranes for x-ray masks produced by laser ablation deposition.* Journal of vacuum science & technology. b, 1991. **9**(6): p. 3254-7.

5. Bourdelle, K.K., *et al.*, *Melting and damage production in silicon carbide under pulsed laser irradiation.* Phys. Stat. Sol., 1990. **121**: p. 399-406.

6. Capano, M.A., *et al.*, *Pulsed laser deposition of silicon carbide at room temperature.* Applied physics letters, 1994. **64**(25): p. 3413-5.

7. Chen, M.Y. and P.T. Murray, *Deposition and characterization of SiC and cordierite thin films grown by pulsed laser evaporation.* Journal of Materials Science, 1990. **25**: p. 4929-32.

8. Fenner, D.B., *et al.*, *Pulsed laser deposition of cadmium telluride, mercury cadmium telluride and beta-silicon carbide thin films on silicon.* Mat. Res. Soc. Symp. Proc., 1992. **268(Materials Modification by Energetic Atoms and Ions)**: p. 235-40.

9. Greenwood, P.F., G.D. Willett, and M.A. Wilson, *Mixed silicon carbide clusters studied by laser ablation Fourier transform ICR mass spectrometry.* Organic Mass Spectrometry, 1993. **28**: p. 831-40.

10. Jean, A., *et al.*, *Biaxial Young's modulus of silicon carbide thin films.* Applied physics letters, 1993. **62**(18): p. 2200-2.

11. Katayama, S., N. Fushiya, and A. Matsunawa, *Laser Physical Vapor Deposition of Si3N4 and SiC, and Film Formation Mechanism.* Transactions of JWRI, 1994. **23**(2): p. 181-9.

12. Martin-Gago, J.A., *et al.*, *Electron loss spectroscopy study of the growth by laser ablation of ultra-thin diamond-like films on Si(100).* Surface Science Letters, 1992. **260**: p. L17-23.

13. Mizunami, T., N. Toyama, and T. Uemura, *Optical emission spectroscopy of ArF-laser-irradiated disilane-acetylene mixtures for 3C-SiC epitaxial growth.* Journal of Applied Physics, 1993. **73**(4): p. 2024-6.

14. Noda, T., *et al.*, *Formation of polycrystalline SiC film by excimer-laser chemical vapour deposition.* Journal of materials science letters, 1992. **11**: p. 477-8.

15.   Noda, T., *et al.*, *Microstructure and growth of SiC film by excimer laser chemical vapour deposition at low temperatures.* Journal of Materials Science, 1993. **28**: p. 2763-8.

16.   Pehrsson, P.E. and R. Kaplan, *Excimer laser cleaing, annealing, and ablation of Beta-SiC.* Journal of materials research, 1989. **4**(6): p. 1480-90.

17.   Rimai, L., *et al.*, *Preparation of Oriented Silicon Carbide Films by Laser Ablation of Ceramic Silicon Carbide Targets.* Applied physics letters, 1991. **59**(18): p. 2266-8.

18.   Rimai, L., *et al.*, *Pulsed layer deposition of SiC films on fused silica and sapphire substrates.* Journal of Applied Physics, 1993. **73**: p. 8242-9.

19.   Rimai, L., *et al.*, *Deposition of thin films of silicon carbide on fused quartz and on sapphire by laser ablation of ceramic silicon carbide targets.* Mat. Res. Soc. Symp. Proc., 1993. **285(Laser Ablation in Materials Processing)**: p. 495-500.

20.   Rimai, L., *et al.*, *Preparation of crystallographically aligned layers of silicon carbide by pulsed laser deposition of carbon onto Si wafers.* Applied physics letters, 1994. **65**(17): p. 2171-3.

21.   Rimai, L., *et al.*, *Deposition of epitaxially oriented films of cubic silicon carbide on silicon by laser ablation: Microstructure of the silicon-silicon- carbide interface.* Journal of Applied Physics, 1995. **77**(12): p. 6601-8.

22.   Scholz, M., W. FuBeta, and K.-L. Kompa, *Chemical vapor deposition of silicon carbide powders using pulsed CO2 lasers.* Advanced materials, 1993. **5**(1): p. 38-40.

23.   Stan, M.A., *et al.*, *Growth of 2H-SiC on 6H-SiC by pulsed laser ablation.* Applied physics letters, 1994. **64**(20): p. 2667-9.

24.   Suzuki, H., H. Araki, and T. Noda, *Effect of Incident Direction of ArF Laser to Graphite Substrates on the Formation of Photo-Chemical Vapor Deposition SiC Film.* Japanese journal of applied physics, Part 1, R, 1993. **32**(8): p. 3566-71.

25.   Suzuki, H., H. Araki, and T. Noda, *Microstructure of SiC thin films produced on graphite by excimer-laser chemical vapour deposition.* Journal of materials science letters, 1994. **13**: p. 49-52.

26.   Tench, R.J., *et al.*, *Clusters formed in laser-induced ablation of Si, SiC, Pt, UO2 and evaporation of UO2 observed by laser ionization time-of-flight mass spectrometry and scanning tunneling microscopy.* Journal of vacuum science & technology. b, 1991. **9**(9): p. 820-4.

27.   Zehnder, T., A. Blatter, and A. Bachli, *SiC thin films prepared by pulsed excimer laser deposition.* Thin solid films, 1994. **241**: p. 138-41.

28.   Zehnder, T., *et al.*, *Tribological properties of laser deposited SiC coatings.* Mat. Res. Soc. Symp. Proc., 1994. **343(Polycrystalline Thin Films: Structure, Texture, Properties and Applications)**: p. 621-6.

Table 1: Summary of SiC PLD growth conditions.

Sample #	Run Date	SUBSTRATE DEPOSITION					PLD CONDITIONS				SiC Thickness		Remarks
		Substrate Type	Cleaning	Carboniz	SCB	SiC thick	Energy(mJ)	Rep-rate(Hz)	Time(min)	Temp(°C)	Total(μm)	Per Pulse(μm)	
1	8/7/95	Si (100)	RCA				240	10	30	938[1]	0.5-1.0	4.2×10^{-5}	No XRD
2	8/8/95	Si (100)	RCA				210	20	60	938[1]	1.0-3.0	2.8×10^{-5}	No XRD
3	8/9/95	Si (100)	RCA				---	10	60	850[2]			Off-center
4	8/10/95	Si (100)	RCA				---	10	60	848[2]	0.8-1.6	3.3×10^{-6}	
5	8/17/95	Si (111)	RCA	AP-1300	LP-900	0.4μm	200	10	60	850[2]	0.7-0.8	2.0×10^{-5}	
6	8/22/95	Si (111)	RCA	AP-1300	LP-900	0.4μm	230	5	60	823[2]	0.3-1.0	4.0×10^{-5}	
7	8/23/95	Si (111)	RCA	AP-1300	LP-900	0.4μm	230	5	60	938[2]	0.5-0.9	4.0×10^{-5}	
8	8/25/95	Si (111)	RCA	AP-1300	LP-900	0.4μm	230	1	90	930[2]	0.4-3.6	6.7×10^{-5}	
9	8/29/95	6H-SiC[3]	RCA				230	1	90	924[2]	0.3-0.4	6.7×10^{-5}	
10	8/30/95	6H-SiC[3]	RCA				230	1	90	1191[2]	0.2-0.3	3.3×10^{-5}	
11	8/31/95	15R-SiC[4]					230	1	90	1196[2]	0.1-0.2	2.8×10^{-5}	

1 Measured with pyrometer aimed at block
2 Measured with pyrometer aimed at substrate
3 J23-#2 Westinghouse:WPAFB
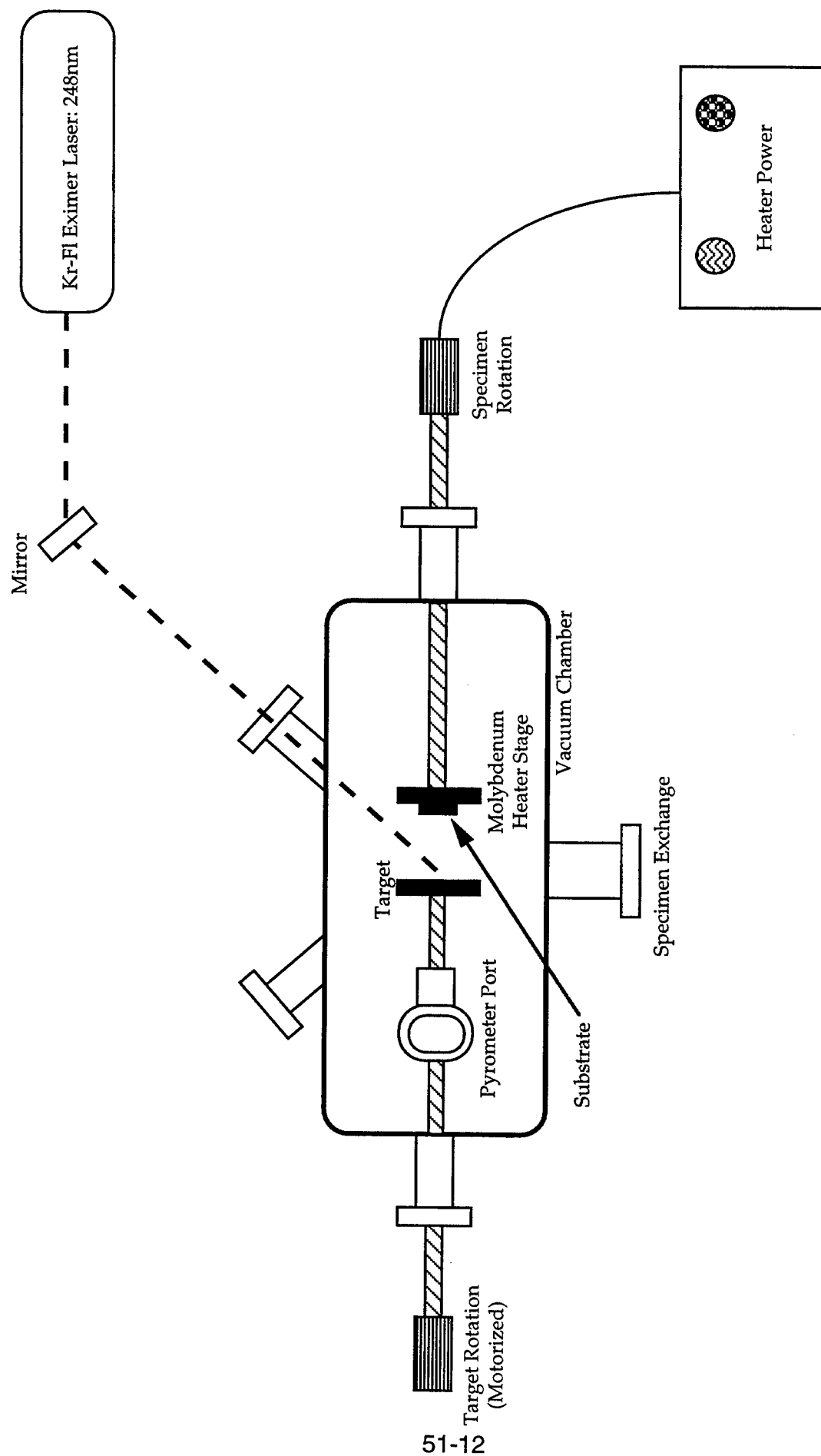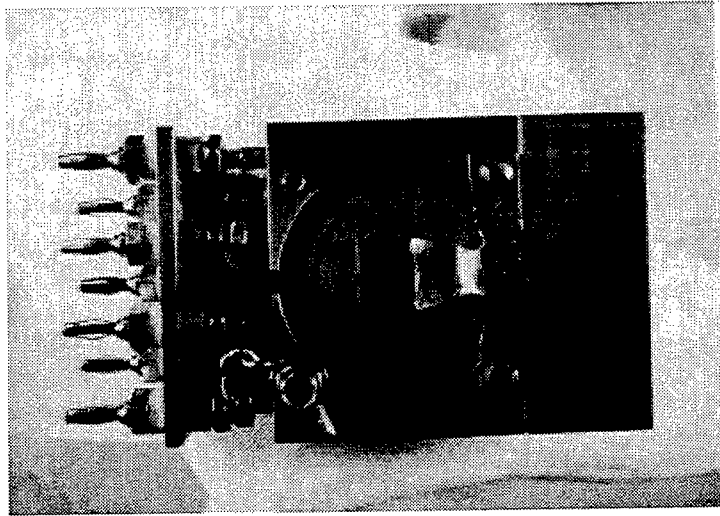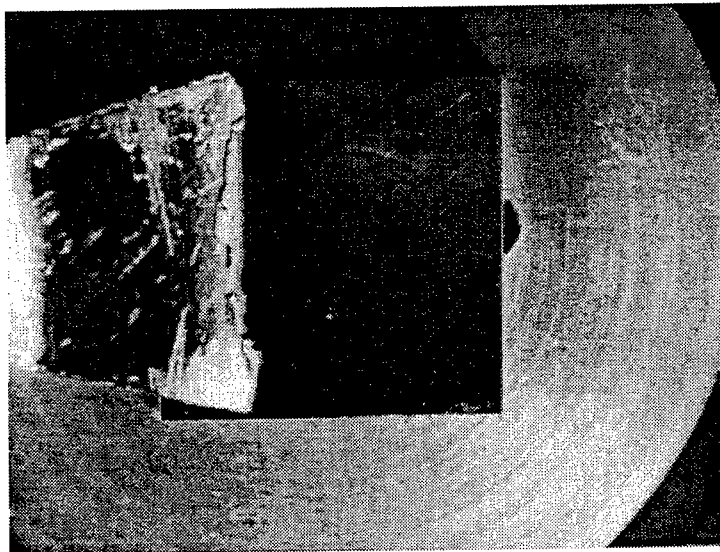4 Lely crystal grown in Kiev, Ukraine

51-11

Kr-Fl Eximer Laser: 248nm

Mirror

Heater Power

Specimen Rotation

Target

Molybdenum Heater Stage

Vacuum Chamber

Specimen Exchange

Pyrometer Port

Substrate

Target Rotation (Motorized)

Figure 1: Pulsed Laser Deposition System

51-12

(a)

(b)

Figure 2: (a) Photo of high temperature heater stage with substrate and (b) close-up of substrate holder with sample.
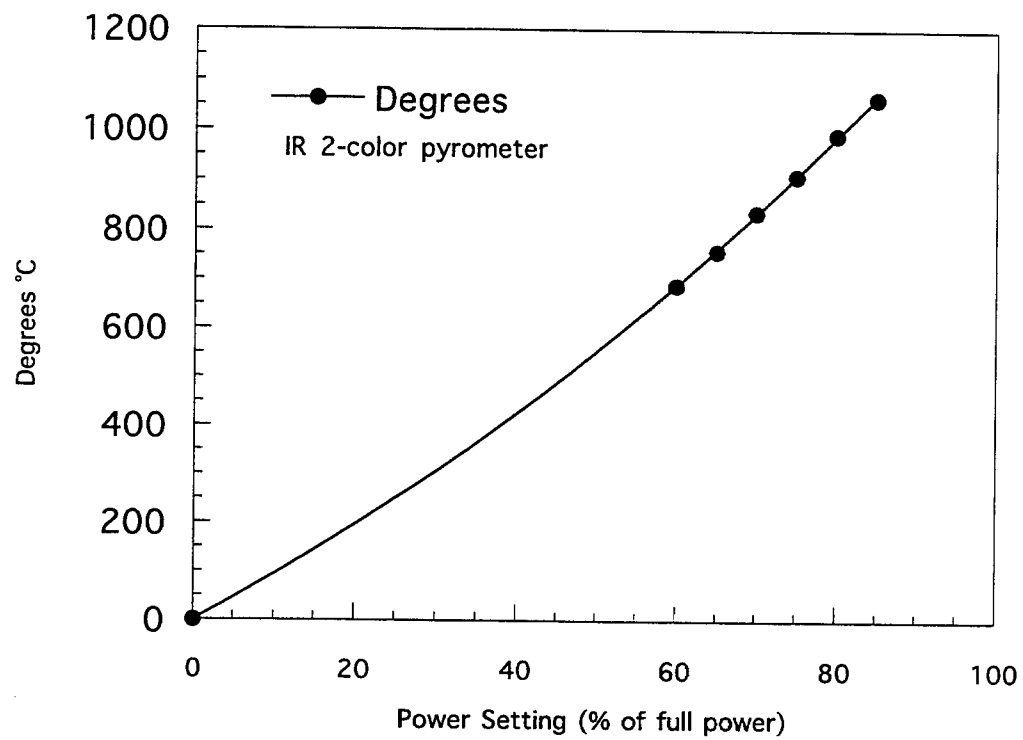
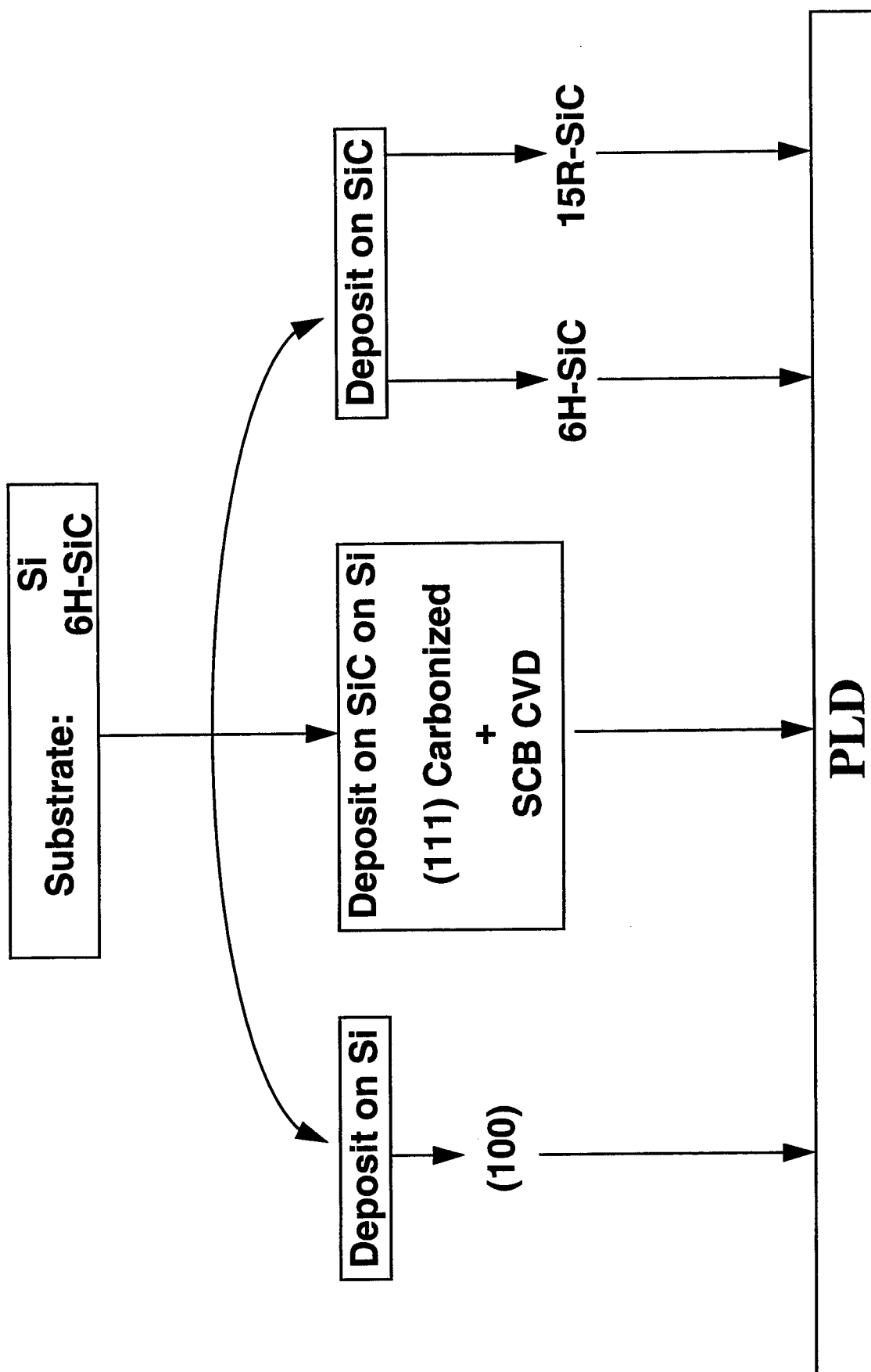**Figure 3:** Molybdenum stage temperature vs. input power

51-14

Pulsed Laser Deposition

Substrate:
Si
6H-SiC

Deposit on Si
(100)

Deposit on SiC on Si
(111) Carbonized
+
SCB CVD

Deposit on SiC
6H-SiC
15R-SiC

PLD

Figure 4: Organization of PLD experiments
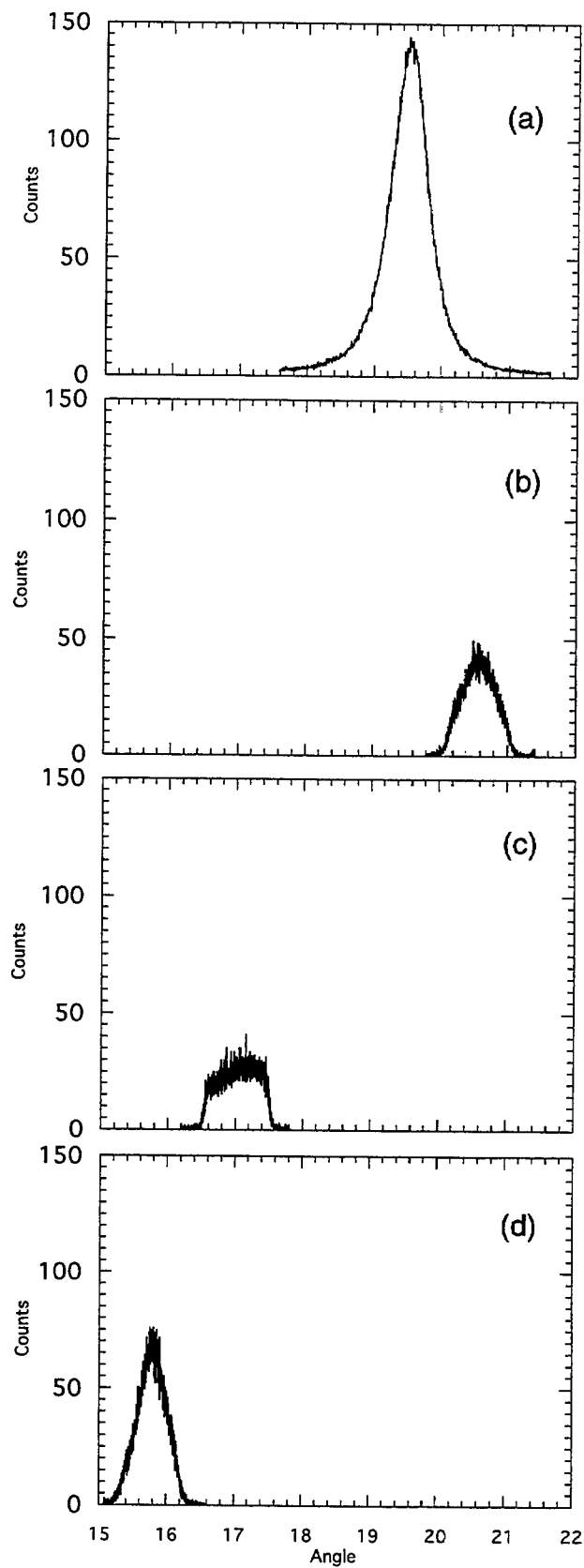
Figure 5: X-ray data from PLD runs on SiC on (111) Si:
(a) Substrate type A1 consisting of (111) carbonized with
low pressure 3C-SiC growth before PLD deposition
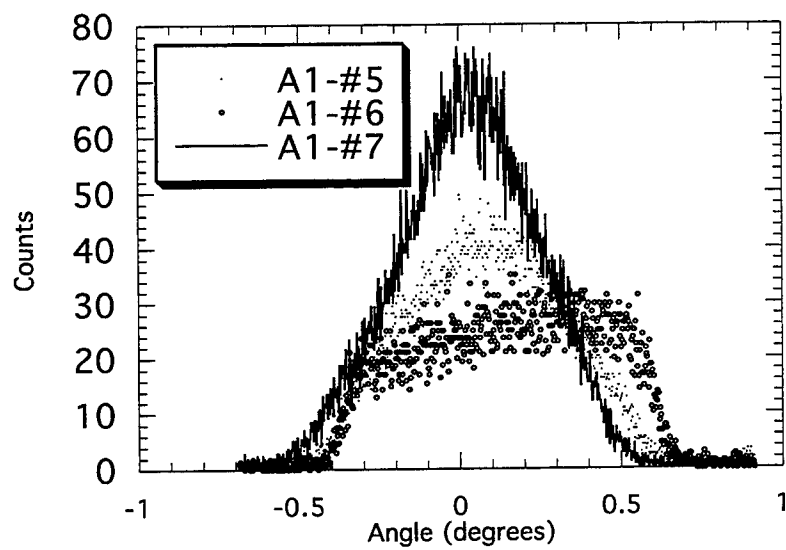(b) Run #5 on A1-(111) Si
(c) Run #6 on A1-(111) Si
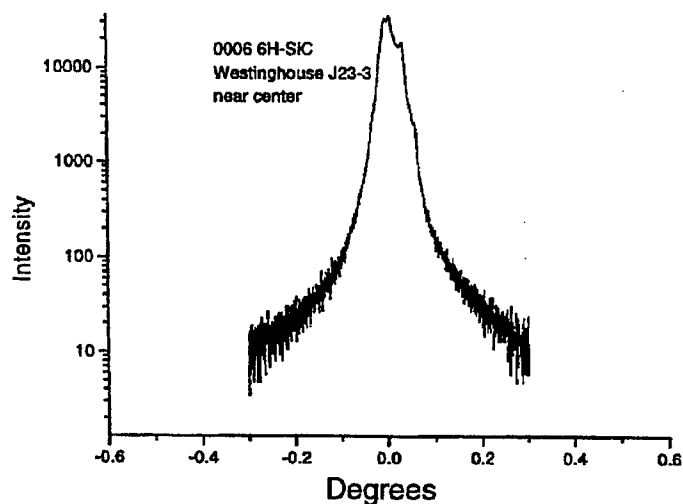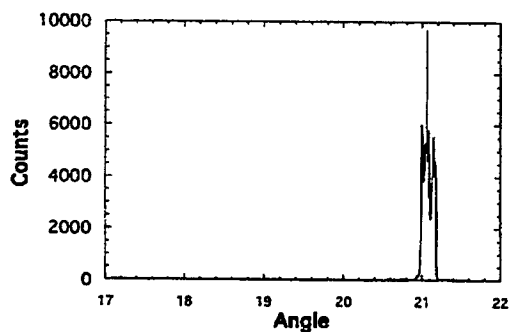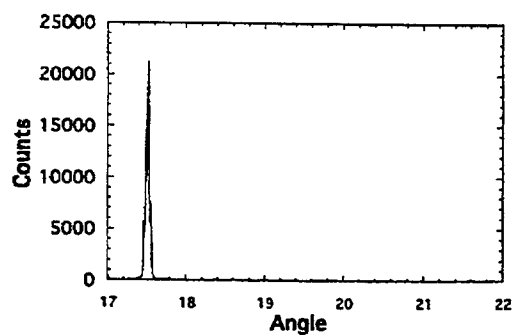(d) Run #7 on A1-(111) Si
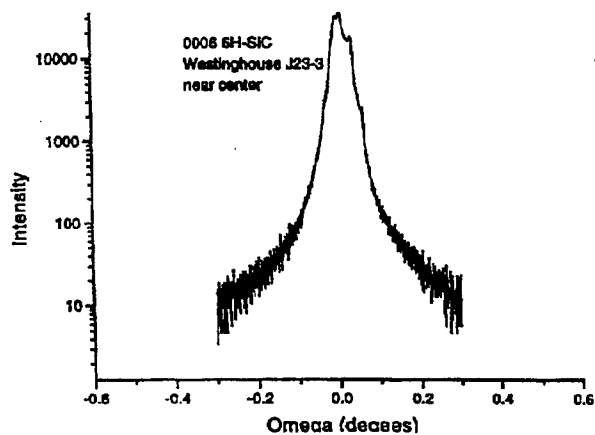
Figure 6: Runs 5-7 on (111) Si substrate

(a) XRD of 6H-SiC substrate grown by Westinghouse
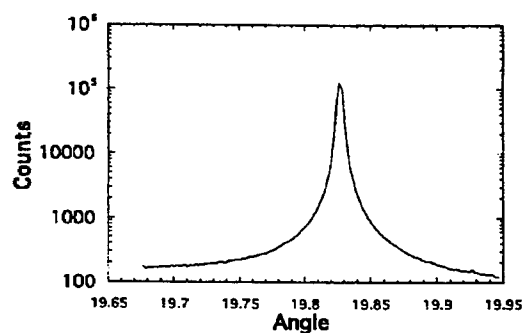before PLD deposition runs #10 &11.

(b) Run #9 on 6H-SiC

(c) Run #10 on 6H-SiC

(d) Initial XRD of Lely-grown 15R-SiC

(e) Run #11 on 15R-SiC (Lely)

Figure 7: X-ray data from PLD growth runs on SiC substrates

51-18

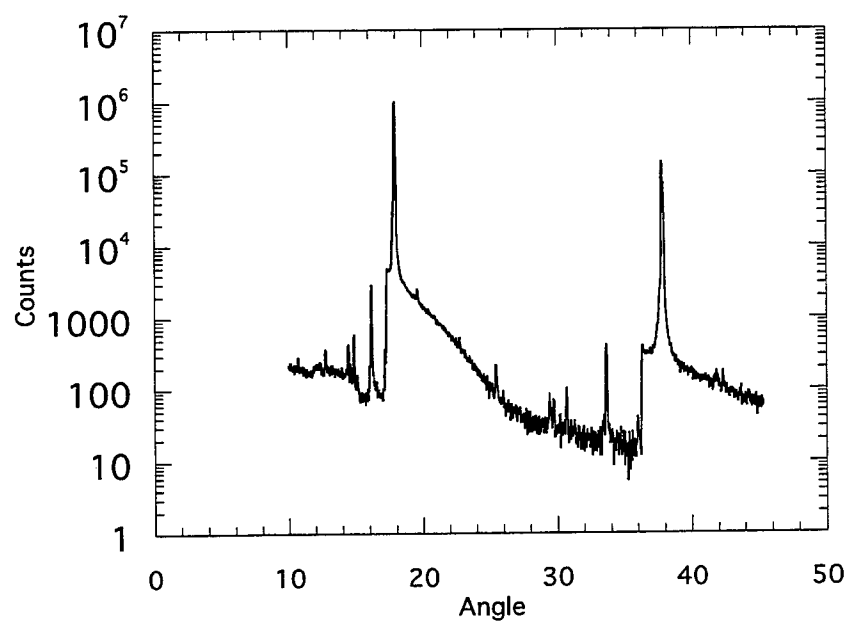**Figure 8:** Large angular scan of PLD on Lely-grown 15R-SiC substrate

ADVANCED PROCESSING TECHNIQUES FOR RESTORATION AND SUPERRESOLUTION OF IMAGERY IN MULTISPECTRAL SEEKER ENVIRONMENTS

Malur K. Sundareshan
Professor of Electrical and Computer Engineering
University of Arizona
Tucson, AZ 85721

Final Report for:
Summer Faculty Research Program
Wright Laboratory Armament Directorate

August 1995

ADVANCED PROCESSING TECHNIQUES FOR RESTORATION AND SUPERRESOLUTION OF IMAGERY IN MULTISPECTRAL SEEKER ENVIRONMENTS

Malur K. Sundareshan
Professor of Electrical and Computer Engineering
University of Arizona
Tucson, AZ 85721

ABSTRACT

This report summarizes the results of a study undertaken to investigate the potential for restoration and superresolution processing of images in multispectral seeker environments for facilitating smart munition guidance. Despite the advances being made in sensor technology, the inherent problems associated with diffraction limited imaging impose limitations on the resolution of acquired imagery thus necessitating some form of post-processing to achieve resolution improvements needed for reliable target detection , classification and aimpoint selection. A mathematical formulation of the restoration and superresolution problems from a frequency spectrum reconstruction viewpoint is given to identify certain critical factors that affect the superresolvability of given image data. Processing characteristics of a few promising approaches for developing systematic algorithms that can be tailored to process image data collected from different types of sensors in multispectral missile seekers are outlined. A brief discussion of the computational requirements and optimization of processing architectures for supporting the implementation of algorithms is given. The role of superresolution processing to facilitate sensor fusion is also outlined. The principal focus of this study is to underscore the importance of optimally tailored restoration and superresolution algorithms for the individual sensor type and operating conditions.

ADVANCED PROCESSING TECHNIQUES FOR RESTORATION AND SUPERRESOLUTION OF IMAGERY IN MULTISPECTRAL SEEKER ENVIRONMENTS

Malur K. Sundareshan

1. INTRODUCTION

Advanced tactical missions capable of efficiently executing diverse mission functions that may include surveillance, threat warning, fire control and countermeasures can not rely on sensors operating solely in one sector of the electromagnetic spectrum. Indeed, a combination of sensors operating over different frequency ranges in a common aperture package would be desirable for facilitating an integrated multi-sensor approach to fulfill diverse mission requirements such as reliable target detection, classification , interleaved acquisition, track and engage modes, and precision kill. While it is evident that deployment of complementary sensors in a missile seeker would enhance the detection, classification and track maintenance performance in addition to providing increased fault tolerance and greater immunity to countermeasures (such as jamming), it also needs implementation of advanced signal processing functions in order to ensure realization of these possible performance benefits.

Notwithstanding the number and diversity of sensors being deployed, one is invariably confronted with problems stemming from the inherent physical limitations of individual sensors. A significant problem is the poor resolution of the collected images which could severely undermine the mission goals. This in turn stems mainly from deployable antenna size limitations (which preclude simply increasing the physical aperture of sensors to gain high image resolution) and the consequent diffraction limits on the achievable resolution. In addition to this problem one has to contend with the degradation effects due to atmospheric conditions and the noise level in the sensing operation. It may be noted that the wavelength of a synthetic aperture radar (SAR) operating at 1 GHz is about 1 inch long and one needs an antenna as big as 40 ft wide in order to achieve a resolution requirement of being able to distinguish points in a scene separated by about 1 meter at a distance of 1 Km [1]. Passive millimeter-wave (PMMW) sensing offers superior adverse weather capabilities (over infra-red (IR) sensors, for instance) due to easy penetration through fog, dust, smoke, etc. However, PMMW image acquisition sensors suffer from poor angular resolution. It is well documented that the angular resolution achievable by a 94 GHz system with a 1 ft diameter antenna is only about 10 mrad, which translates into a spatial resolution of about 10 meters at a distance of 1 Km. Some recent studies [2] have also established that for ensuring reasonably adequate angular resolution (typically of the order of 4 mrad), a 94 GHz PMMW imaging system with a sensor depression angle of $60^0 - 80^0$ needs to be confined to very low operational altitudes (of the order of 75-100 meters) which puts inordinate demands on the guidance schemes to facilitate such requirements. Similar resolution limitations and the consequent requirements on operational conditions (some of which may be clearly impossible to satisfy for tactical missions with reliability and survivability constraints) exist for the other types of sensing modalities as well.

Typical seeker antenna patterns are of a " low-pass" filtering nature due to the finite size of the antenna or lens that makes up the imaging system and the consequent imposition of the underlying diffraction limits. Hence the image recorded at the output of the imaging system is a low-pass filtered version of the original scene. The portions of the scene that are lost by the imaging system are the fine details (high frequency spectral components) that

accurately describe the objects in the scene, which also are critical for reliable detection and classification of targets of interest in the scene. Hence some form of image processing to restore the details and improve the resolution of the image will invariably be needed. Traditional image restoration procedures (based on deconvolution and inverse filtering approaches) attempt mainly at reconstruction of the passband and possibly elimination of effects of additive noise components. These hence have only limited resolution enhancement capabilities. Greater resolution improvements can only be achieved through a class of more sophisticated algorithms, called superresolution algorithms, which provide not only passband reconstruction but also some degree of **spectral extrapolation,** thus enabling to restore the high frequency spatial amplitude variations relating to the spatial resolution of the sensor and lost through the filtering effcts of the seeker antenna pattern. A tactful utilization of the imaging instrument's characteristics and any *a priori* knowledge of the features of the target together with an appropriately crafted nonlinear processing scheme is what gives the capability to these algorithms for superresolving the input image by extrapolating beyond the passband range and thus extending the image bandwidth beyond the diffraction limit of the imaging sensor.

For application in missile seeker environments, it must be emphasized that superresolution is a post-processing operation applied to the acquired imagery and consequently is much less expensive compared to improving the imaging system for desired resolution. As an example, it may be noted that for visual imagery acquired from space-borne platforms, some studies indicate that the cost of camera payload increases as the inverse 2.5 power of the resolution. Hence a possible two-fold improvement in resolution by superresolution processing in this application roughly translates into a reduction in the cost of the sensor by more than 5 times. Similar relations also exist for sensors operating in the other spectral ranges (due to the relation between resolution and antenna size), confirming the cost effectiveness of employing superresolution algorithms. The principal goal of superresolution processing in multispectral seekers is hence to obtain an image of a target of interest (such as tactical mobile and extended area high value targets) via post-processing that is equivalent to one acquired through a more expensive larger aperture sensor.

Most of the recent work in the development of image restoration and superresolution algorithms has been motivated by applications in Radioastronomy and Medical Imaging. While this work has given rise to some mathematically elegant approaches and powerful algorithms, a certain degree of care should be exercised in adapting these approaches and algorithms to the missile seeker environment. This is due to the convergence problems often encountered by iterative schemes and the specific statistical models representing the scenarios facilitating their development. For example, a slowly converging algorithm that ultimately guarantees the best resolution in the processed image may pose no implementational problems in Radioastronomy; however, it could be entirely unrealistic for implementation in an autonomous unmanned tactical system that must operate fast enough to track target motion. Even when use within the same application area is contemplated, the selected algorithm should take into account the basic properties of the signals being processed. For instance, a maximum likelihood estimation algorithm developed with Poisson distribution models may be capable of ensuring satisfactory resolution improvements in processing optical images. It may however become inappropriate for PMMW images or IR images due to the infeasibility of such a model to accurately portray the reflectivity characteristics from targets and clutter at these frequencies. Hence a careful tailoring of the processing algorithm for each sensor supporting

the multispectral seeker is of critical importance in order to realize the possible performance benefits from superresolution processing which include better false target rejection, improved automatic target recognition and aimpoint selection.

This report will discuss the processing characteristics and implementational requirements of a few selected superresolution approaches for multispectral seeker environments. In Section 2, we shall briefly describe a mathematical formulation of the image restoration and superresolution problems from a spectrum reconstruction viewpoint and identify certain factors that would impose limitations on the superresolvability of given image data. Some promising approaches for developing specific superresolution algorithms will be outlined in Section 3. Section 4 will discuss the requirements and optimization of processing architectures for supporting implementation of algorithms to process image data from missile seekers. The role of superresolution processing to facilitate sensor fusion and multispectral enhancement will also be outlined in this section. Finally in Section 5, the principal conclusions from this study will be summarized.

2. RESTORATION AND SUPERRESOLUTION OF SENSOR OUTPUTS

2.1 Image Formation Process (Observation Model)

Every systematic image processing study (including image restoration and superresolution processing) will start with an appropriate mathematical model characterizing the process of image formation by the sensor employed, which is termed an " observation model". Irrespective of the type of sensor actually used, a commonly used observation model takes the form

$$g = s (Hf) + n \qquad\qquad (1)$$

where f denotes the object being sensed, g its image and H denotes the operator that models the filtering process including any associated degradations (such as due to small aperture size of the sensor) and blur phenomena (caused by atmospheric effects, motion of object or the sensor, or out of focus operations, etc.). n denotes the additive random noise in the sensing process, which includes both the receiver noise and any quantization noise. The response of the image recording sensor to the intensity of input signal (light, radar, etc.) is represented by the memoryless mapping $s(.)$, which is in general nonlinear.

For the sake of precision, let us consider the image to be obtained from an incoherent sensor. We will also assume that the image to be processed consists of $M \times M$ equally spaced grey level pixels, obtained through a sampling of the image field at a rate that satisfies the Nyquist criterion. Furthermore, for mathematical tractability we will make the commonly used assumptions which include: (i) space-invariant imaging process, (ii) ignore the nonlinear effects of the sensor, and (iii) approximate the noise process by a zero-mean white Gaussian random field which is independent of the object. With these assumptions, Equation (1) can be rewritten to relate the image intensity value $g(i,j)$ at pixel (i,j) to the object pixel values as

$$g(i,j) = \sum_{(k,l) \in S} h(i-k,j-l)f(k,l) + n(i,j), \qquad i,j = 1,2,....,M. \quad (2)$$

where $h(i,j)$ denotes the point spread function (PSF) of the sensor.

For an image of size $M \times M$, Equation (2) corresponds to a set of M^2 scalar equations specifying the formation of each image pixel. For a further simplified representation [3,4], by a lexicographical ordering of the signals g, f

and n, one can rewrite Equation (2) as resulting from a convolution of two one dimensional vectors $\mathbf{h} = \left[h_1, h_2, ..., h_N\right]^T$ and $\mathbf{f} = \left[f_1, f_2, ..., f_N\right]^T$ as

$$g_i = h_i \otimes f_i + n_i = \sum_{j=1}^{N} h_{i-j} f_j + n_i \quad , \qquad i = 1, 2, ..., N \, , \qquad (3)$$

where $N = M^2$. More compactly, Equation (3) can be rewritten as the vector equation

$$\mathbf{g} = \mathbf{Hf} + \mathbf{n} \qquad (4)$$

where $\mathbf{g}$, $\mathbf{f}$ and $\mathbf{n}$ are vectors of dimension N and $\mathbf{H}$ denotes the PSF block matrix whose elements can be constructed [3,4] from the PSF samples $\left\{h_1, h_2, ..., h_N\right\}$. It should be noted that Equations (3) and (4) represent space-domain models and are equivalent to the frequency-domain model

$$\mathbf{G}(\omega) = \mathbf{H}(\omega)\mathbf{F}(\omega) + \mathbf{N}(\omega) \qquad (5)$$

where ω is the discrete frequency variable and $\mathbf{G}(\omega)$, $\mathbf{F}(\omega)$, $\mathbf{H}(\omega)$ and $\mathbf{N}(\omega)$ are the DFT's of the N-point sequences $\left\{g_i\right\}, \left\{f_i\right\}, \left\{h_i\right\}$ and $\left\{n_i\right\}$ respectively.

2.2 Problem of Interest

For application in missile seeker environments, Equation (3) describes the process of image formation when an unknown object with radiance distribution $\left\{f_i\right\}$ is imaged through a sensor with a shift-invariant PSF $\left\{h_i\right\}$. As noted earlier, practical seeker antenna patterns have a low-pass spectral characteristic and consequently the image obtained is a low-pass filtered version of the object (or scene) being imaged. The problem of interest is then to recover the object, i.e. $\left\{f_i\right\}$, by solving Equation (3) (or Equation (5)). However, since the noise sequence $\left\{n_i\right\}$ will not be known exactly, one will not be able to solve Equation (3) for $\left\{f_i\right\}$ exactly even when $\left\{h_i\right\}$, the PSF of the seeker antenna, is exactly known. One can only hope to obtain an estimate $\left\{\hat{f}_i\right\}$ which is in some sense close to the original $\left\{f_i\right\}$, based on some reasonable assumptions on the noise process $\left\{n_i\right\}$. If a distance measure $\mathbf{J}$ $(\mathbf{g}, \mathbf{f})$ between $\mathbf{g}$ and $\mathbf{f}$ is used as a norm to measure the closeness of the estimate, the problem of interest can be specified concisely as obtaining the estimate $\hat{\mathbf{f}}^T = \left[\hat{f}_1 \ \hat{f}_2 \ .... \ \hat{f}_N\right]$ such that

$$\hat{\mathbf{f}} = \arg \min_f \mathbf{J} (\mathbf{g}, \mathbf{f}) \quad = \arg \min_f \mathbf{J}\left(g_i - \sum_j h_{ij} f_j\right) \qquad (6)$$

An examination of the frequency spectra of the object and the image is useful to see clearly the effect of the seeker antenna. Let us assume that the object is space-limited with spatial extent χ and, without any loss of generality, assume that $\left\{f_i\right\}$ is nonzero only on the interval $\left[-\frac{\chi}{2}, +\frac{\chi}{2}\right]$. This implies that the spectrum $\mathbf{F}(\omega)$ has infinite extent, i.e. the object has infinite bandwidth, and in the discrete frequency domain, the spectral components in $\mathbf{F}(\omega)$ extend all the way to $\frac{\omega_s}{2}$, the folding frequency, as shown in Fig. 1a.

The image spectrum $\mathbf{G}(\omega)$ is a low-pass filtered version of $\mathbf{F}(\omega)$ with the cut-off frequency ω_c determined by the diffraction limit of the sensor. Assuming an ideal low-pass filter characteristic, the shape of $\mathbf{G}(\omega)$ will be as shown in Fig. 1b with the spectral components removed in the interval $\omega_c \leq \omega \leq \omega_s / 2$. The degradations in the image are hence caused by three factors: (1) spectral mixing within the passband $0 \leq \omega \leq \omega_c$ due to the convolution with the PSF of the seeker antenna; (2) spectral attenuation caused by removal of spectral components

outside the passband; and (3) corruption of components in the passband due to the additive noise process $\{n_i\}$.
Perfect image restoration requires compensation for all three factors cited above.
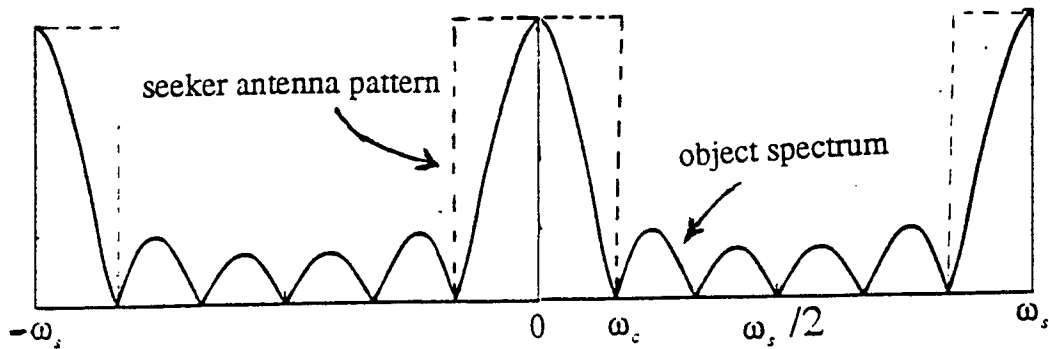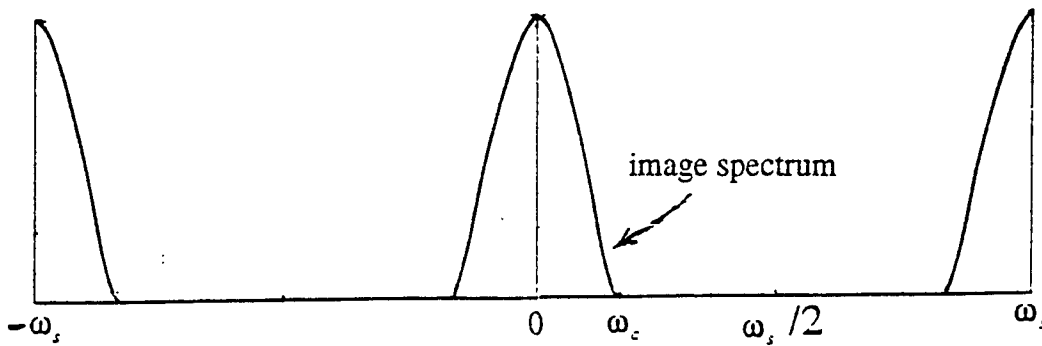

Fig. 1a

Fig. 1b

Traditional image restoration methods attempt mainly at passband reconstruction, i.e. to eliminate the degradations caused by the first and the third factors above. This is achieved by various deconvolution and noise filtering approaches [4,5]. The goal of superresolution is to correct for all three of the above factors, and hence in addition to restoration of spectral components in the passband, extrapolation of the spectrum beyond ω_c is to be attempted. Evidently, the ideal of restoring all lost spectral components may be too ambitious and hence realistically one may have to be content with some spectral extrapolation which facilitates recovering the spectrum in the interval $\omega_c \leq \omega \leq \omega_e$, where $\omega_e < \omega_s / 2$ is an extended frquency limit. It is of interest to note that even if this limited goal is attained, then the effective cut-off frequency is moved from ω_c to ω_e and hence the processed image appears as the image acquired from a higher resolution (more expensive, larger aperture) sensor with this larger cut-off frequency. It should also be noted that since generation of new frequency components not present in the original image is attempted, some form of nonlinear processing becomes essential, since linear signal processing methods can not produce frequencies not present in the input signal.

To illustrate the complexity in solving problems of this type, consider the simplest case when no spectral extrapolation is needed, the PSF $\{h_i\}$ of the sensor is assumed to be known and the noise n_i is ignored. This is the classical deconvolution problem [4,5] of solving the vector equation

$$\mathbf{g} = \mathbf{H}\mathbf{f} \tag{7}$$

for $\mathbf{f}$ given $\mathbf{g}$ and $\mathbf{H}$, and a solution can be attempted in the form of an "inverse filter" given by

$$\hat{\mathbf{f}} = \mathbf{H}^{-1}\mathbf{g}. \tag{8}$$

Unfortunately, there are several problems with this approach. The system of equations given by Equation (7) is often underdetermined which results in $\mathbf{H}^{-1}$ being not defined. Even if $\mathbf{H}^{-1}$ (or a generalized inverse of $\mathbf{H}$) can be

computed, the estimate $\hat{f}$ obtained may be worthless due to the presence of noise that was ignored. Observe from the image formation model given by Equation (4), when the presence of is accounted for,

$$\hat{f} = H^{-1}g - H^{-1}n.$$

It is now clear that H being a low-pass filter, H^{-1} corresponds to a high-pass filter and hence the noise is greatly amplified in the solution estimate. A difficulty of a related nature which also can make the solution given by Equation (8) of limited value is that an exact knowledge of H is needed for computing the solution and even a small uncertainty in the parameters describing the sensor PSF can result in a very large discrepancy in the solution. In other words, the solution given by Equation (8) is not "robust" enough to tolerate these nonideal conditions that may exist in practice making the estimate obtained useless. Finally, the inverse filter solution is a linear operation and provides no extrapolation of spectrum thus lacking any capability for superresolving. It will be seen later that the drawbacks of this solution procedure stem from the fact that no use of any *a priori* knowledge available about the object being restored is made.

2.3   Limits on Superresolution

The idea of recreating the spectral components that are removed by the imaging process and hence are not present in the image available for processing may pose some conceptual difficulties, which may lead one to suspect whether superresolution is indeed possible. Fortunately there exist sound mathematical arguments confirming the possibility of spectral extrapolation. The primary justification comes from the Analytic Continuation Theorem and the property that when an object has finite spatial extent its frequency spectrum is analytic [6]. Due to the property that a finite segment of any analytic function in principle determines the whole function uniquely, it can be readily proved that knowledge of the passband spectrum of the object allows a unique continuation of the spectrum beyond the diffraction limit imposed by the imaging system. It must be emphasized that the limited spatial extent of the object is critical in providing this capability for extrapolation in the frequency domain.

The Gerchberg-Papoulis formalism [8,9] identifies a possible iterative procedure for solving the superresolution problem. This algorithm alternately applies constraints in the space-domain and the frequency-domain in the quest for reconstructing the unknown portion of the frequency spectrum from a limited known portion, viz. the passband of the sensor, when some *a priori* knowledge of the spatial extent of the object is available (the object values within this extent are not known , however). Each iteration of the algorithm is a two-step procedure that can be summarized as under:

STEP 1-- Transform known passband spectrum to space-domain and apply space limit constraint to set object values to zero outside the region where it is known to be space limited.

STEP 2-- Transform result back to frequency-domain and correct the passband back to the known original values.

These two steps are iteratively applied until a satisfactory estimate of the spectral components being reconstructed is obtained. Analytical support for the convergence of the procedure comes from consideration of the energy of the error function (error between the true spectrum and its estimate). The application of constraints in each iteration, once in the space-domain and the other in the frequency-domain, provides the nonlinear operations to generate the new frequencies needed for superresolution.

The arguments given above clearly attest to the possibility of spectral extrapolation beyond the sensor cutoff frequency ω_c. The question then is to determine whether there exist any limits on the bandwidth extension possible, i.e. the extended frequency ω_e, or whether it is possible to reconstruct the entire spectrum (up to the limit imposed by the sampling rate used, viz. $\omega_s / 2$). Evidently, the amount and the quality of bandwidth extension are measures of the performance of a superresolution algorithm. Since passband reconstruction performance is also very important, a useful question to pose is -- given a bound $\beta > 0$ serving as a measure of the tolerable spectral deviation, find the largest ω_e such that $\left\| \hat{F}(\omega) - F(\omega) \right\| \leq \beta$ for all ω in the interval $0 \leq \omega \leq \omega_e$.

The development of limits for signal recovery has been of interest in Communications and Information Sciences for a number of years (Shannon's information transfer limits). In the present context, it is intuitive to expect that the limits on ω_e come from a number of factors, most notably those that affect the precision of the data being processed. Clearly, one of the important factors is the noise level present, i.e. SNR of the image. The role of SNR in affecting the quality of image restoration has long been studied. In fact, many of the available approaches for deconvolution start from recognizing the problem as an "ill-posed problem" (i.e. estimate $\hat{f}$ of the restored image is not a continuous function of image data g and arbitrarily small perturbations in g can lead to arbitrarily large variations in $\hat{f}$) and aatempting to devise procedures (such as regularization [10,11]) for overcoming this problem.

Several other factors can equally influence any limits on ω_e and consequently establishing precise analytical limits for the bandwidth extension possible in all cases seems infeasible. For optical images, by using an information theoretic analysis, Kosarev [12] determined a resolution limit under certain conditions to be

$$r_l = \frac{1}{\omega_c \log_2(1 + SNR)}$$

where r_l is the minimum resolvable distance between two point sources in the object, ω_c is the cut-off frequency of the imaging system and SNR is the image SNR. It may be noted that the quantity on the right-hand-side is fixed for any specific image data and hence any separation greater than the value of r_l can be potentially resolved by superresolution processing. This formula also specifies the role of ω_c in obtaining resolution limits. Note that as ω_c increases, r_l decreases (i.e. resolution increases) which agrees well with intuition.

Another term that affects any limits on ω_e is the sampling rate used. Clearly, higher the value of ω_s, larger will be the width of the frequency band $\omega_c \leq \omega \leq \omega_s / 2$ for any imaging system, and greater is the potential for resolution improvement. In fact, many practical superresolution algorithms use interpolation or other upsampling procedures [13] appropriately during execution of the algorithm to obtain the effect of larger ω_s.

One factor that is not readily apparent in regard to its relation to resolution limits is the spatial extent of the object. Very recently, Sementilli et.al. [14] have derived an approximate bound on bandwidth extrapolation for optical images which involves ω_c, ω_s and the object extent. This bound not only establishes that the potential for superresolution increases with decreasing spatial extent of objects being imaged, but also confirms the intuitive feeling that restoration results often demonstrated for two-point source objects do not necessarily translate into corresponding performances in the case of spatially extended objects.

It should be emphasized that while analytical limits such as the ones discussed above are useful for determining potential benefits of superresolution processing of given image data, one does not have to be necessarily discouraged if the data to be processed does not meet these limits. Observe that only the characteristics of the image data g and of the imaging system, and only a limited information on the object (viz., spatial extent) are used in developing these limits. If any additional knowledge of the object being restored is available, one may attempt to use this information for possible spectral extrapolation and hence improved resolution. How to utilize this information is at the heart of a well-tailored superresolution algorithm.

3. APPROACHES TO SUPERRESOLUTION PROCESSING

3.1 General Requirements and Use of *a priori* knowledge

As noted in the last section, due to the ill-posed nature of the inverse filtering problem underlying image restoration and superresolution objectives, it is necessary to have some *a priori* information about the ideal solution, i.e. the object f being restored from its image g. In algorithm development, this information is used in defining appropriate constraints on the solution and/or in defining a criterion for the "goodness" of the solution. It may be recalled that the use of such constraints is fundamental in the application of Gerchberg-Papoulis formalism and is in fact the basis for the nonlinear processing so necessary for superresolution.

The specific *a priori* knowledge that can be used evidently depends on the specific application. For applications in astronomy, it could come in the form of some known facts about the spectral differences of the objects one is looking for (for instance, a double star as opposed to a star cluster). In medical imaging and in military applications, it could come from the geometrical features of the object (target shape, for instance). For radar and MMW imagery, one could use the fundamental knowledge that the reflectivity of any point on the ground can not be negative. In addition to the nonnegativity constraint, a space constraint resulting from the known space-domain limits on the object of interest could be used. Other typically available constraints include level constraints (which impose upper and lower bounds on the intensity estimates $\hat{f}_j$), smoothness constraints (which force neighboring pixels in the restored image to have similar intensity values) and edge-preserving constraints. More complicated constraints are possible, but in general they result tuning the algorithms to specific classes of targets.

Varying by the extent to which *a priori* knowledge can be incorporated in algorithm development, there have been introduced into the literature a large number of image restoration approaches and algorithms too vast to describe or reference here. One may refer to some recent survey papers [17,18] for a review of the extensive activity on this topic. In this section, we shall only briefly cite a few of the approaches that have received some interest in the context of superresolution capabilities, i.e. those that provide possible spectral extrapolation. It should be noted clearly that not all image restoration methods provide the capability for superresolving. In fact, a majority of existing schemes may perform decent passband restoration, but provide no bandwidth extension at all.

The various approaches in general attempt to code the *a priori* knowledge to be used by specifying an object model or a set of constraint functions, and further employ an appropriate optimization criterion to guide in the search for the best estimate of the object. A convenient way of classifying the resulting algorithms is into **iterative** and **noniterative** (or direct) schemes. Noniterative approaches generally attempt to implement an inverse filtering

operation (without actually performing the computation of the inverse of the PSF matrix **H**, however) and have poor noise characteristics. All required computations and any possible use of constraint functions are applied in one step. In contrast, iterative methods apply the constraints in a distributed fashion as the solution progresses and hence the computations at each iteration will be generally less intensive than the single-step computation of noniterative approaches. Some additional advantages of iterative techniques are that, (1) they are more robust to errors in the modeling of the image formation process (uncertainties in the elements of the PSF matrix **H**, for instance), (2) the solution process can be better monitored as it progresses, (3) constraints can be utilized to better control the effects of noise (and possibly clutter), and (4) can be tailored to offset sensor nonlinearities. The disadvantages of these methods generally are, (1) increased computation time,and (2) need for proving convergence of the iterative scheme (in fact, for some algorithms this could be impossible). Despite these disadvantages, iterative methods are generally the preferred approach due to their numerous advantages and also since the iteration can be terminated once a solution of a reasonable quality is achieved.

In the following we shall very briefly outline a few algorithms that have received some attention in superresolution literature. Due to page limitations, performance details of these algorithms and a comparative evaluation of them will not be given.

3.2   Noniterative (Direct) Algorithms

One of the more well known algorithms in this category was given by Gleed and Lettington [19] using a regularized pseudo-inverse computation approach. Starting with the space-domain image formation model given by Equation (4), Gleed and Lettington note that evaluating the solution as

$$\hat{f} = H^{-1}g - H^{-1}n \qquad (9)$$

provides a poor quality estimate due to the noise amplification caused by H^{-1} (in turn due to some eigenvalues of H becoming too small). To overcome this difficulty, they propose to modify the estimate by first diagonalizing the H matrix through the transformation

$$M^{T}HM = \Lambda$$

where M is the modal matrix of H and Λ is a diagonal matrix with the eigenvalues of H along the diagonal [20]. The object estimate $\hat{f}$ is then obtained as

$$\hat{f} = H_{\text{mod}}^{-1}g - H_{\text{mod}}^{-1}n \qquad (10)$$

where H_{mod}^{-1} is computed as

$$H_{\text{mod}}^{-1} = M\left[\Lambda + \mu_{1}\Lambda^{-1}\right]^{-1}M^{T} . \qquad (11)$$

$\mu_{1} \geq 0$ is a scalar parameter to be selected appropriately based on the noise present n. Observing however that this solution changes the PSF of the imaging system (from H to H_{mod}), Gleed and Lettington propose a "regularization" operation by constructing the matrix R as

$$R = H_{\text{mod}}^{-1}H + \mu_{2}\left(H_{\text{mod}}^{-1}H\right)^{-1} \qquad (12)$$

and obtaining the final estimate as

$$\hat{f} = R^{-1}H_{\text{mod}}^{-1}g . \qquad (13)$$

In Equation (12), μ_{2} is another user selected parameter satisfying the condition $\mu_{2} \leq \mu_{1}$. Gleed and Lettington

[19] report getting satisfactory resolution improvements in processing various images including PMMW imagery. The exact extent of spectral extrapolation obtained by this method is however not clear. Furthermore, the selection of scalars μ_1 and μ_2 is rather ad hoc.

Another recent algorithm of this type that employs a least squares approach to solve the inverse filtering problem but avoids explicit matrix inversion is given by Walsh and Delaney [7]. These two procedures are quite simple and give a noniterative one-step procedure that involves matrix computations. However, both algorithms in essence attempt to implement an inverse filter (i.e. H^{-1}), and hence the noise handling and bandwidth extension properties are not clear in the general cases. Furthermore, the exact knowledge of PSF matrix H is very critical since no *a priori* knowledge of f is utilized in the solution process.

3.3   Iterative Deconvolution Procedures

An iterative approximation to the inverse filtering operation without explicitly inverting the PSF matrix H can be obtained and this is the basis for the class of algorithms that are popularly referred to as Iterative Deconvolution Procedures. Like the direct procedures discussed in the last section, these also employ a deterministic framework (by setting noise $n = 0$) for the algorithm development and attempt to account for presence of noise later. these procedures also have a long history going back to Van Cittert's iteration which is obtained by using the identity

$$f = f + \beta(g - Hf) \qquad (14)$$

which must hold for all values of the "gain" parameter β if the imaging equation (4) is satisfied. One can then use the method of successive approximations to set up the iteration

$$\hat{f}_{k+1} = \hat{f}_k + \beta\left(g - H\hat{f}_k\right) \quad ; \quad \hat{f}_0 = \beta g \qquad (15)$$

where $\hat{f}_k$ denotes the object estimate at the k-th iteration step. While convergence of a procedure such as this needs to be established separately, the major benefits of being able to enforce any constraint from the *a priori* knowledge about f at each iteration to correct for any dicrepancies are readily apparent.

A more recently developed algorithm of this type is the Constrained Iterative Deconvolution (CID) procedure due to Schafer et.al. [22]. This algorithm implements an approximation to the frequency-domain inverse filtering operation

$$\hat{F}(\omega) = H^{-1}(\omega)G(\omega) \qquad (16)$$

based on a series expansion of the inverse filter $H^{-1}(\omega) = \beta\sum_{i=0}^{\infty}\left(1 - \beta H(\omega)\right)^i$ which leads to an iterative algorithm that approximately builds the series one term at a time as

$$\hat{F}_{k+1}(\omega) = \hat{F}_k(\omega) + \beta\left(G(\omega) - H(\omega)\hat{F}_k(\omega)\right) \quad , \quad \hat{F}_0 = G \qquad (17)$$

The gain parameter β needs to be appropriately selected to ensure convergence of the scheme, a condition for which is given by $\|\beta H(\omega)\| \le 1$.

It should be emphasized that this iterative approach merely implements the inverse filter which is linear and hence can not provide any spectral extrapolation. For generating the needed extra frequencies, constraints are enforced at each iteration to modify the algorithm into

$$\hat{F}_{k+1}(\omega) = \aleph\left(\hat{F}_k(\omega) + \left(G(\omega) - H\hat{F}_k(\omega)\right)\right) \qquad (18)$$

where $\aleph(.)$ denotes the constraint function. A typical constraint function is one that enforces the nonnegativity condition on the space-domain estimate $\hat{f}$, which can be implemented by simply truncating the negative portion of the signal. The use of constraints in this manner provides a method to incorporate *a priori* knowledge and hence permits expansion of bandwidth.

Attempts at improving the speed of the algorithm (by a factorization of the series expansion for $H^{-1}(\omega)$) and at using this approach for processing radar images have been made by Richards [23]. Despite the successes reported, it appears that the CID algorithm and its modifications have some weaknesses in processing images with extended targets. Some recent studies have also indicated that these algorithms perform rather poorly in noisy environments (even small amounts of additive noise tend to generate spurious targets, as observed by Ding [24].

3.4 Bayesian Methods

Unlike the previously discussed approaches, these methods employ a stochastic framework. Bayesian methods for restoration of degraded images generally involve using Maximum Likelihood (ML) or Maximum a Posteriori (MAP) techniques to obtain the object estimate $\hat{f}$ from the image g by assuming certain probability distributions for f and g. Fundamental to this approach is the Bayes rule which relates these probability distributions in the form

$$p(f/g) = \frac{p(g/f)p(f)}{p(g)} \qquad (19)$$

where $p(f/g)$, the posterior density function, summarizes the full state of knowledge concerning the imaging process. $p(g/f)$ denotes the probability distribution of observed data conditioned upon the object f and is called the "likelihood", $p(f)$ is the prior density of f (called the "prior") and $p(g)$ is the probabilty that image g will be observed. It should be noted that Equation (19) provides a mechanism for intelligently using *a priori* knowledge on f by specifying an appropriate statistical model for the prior $p(f)$ and combine this with the available data (viz., the image g). It may also be noted that the probabilities of the various quantities are functions of continuous parameters , namely the image pixels.

The likelihood $p(g/f)$ is completely determined by the imaging process model (viz., Equation (4)) and the noise distribution. Thus, in a sense, Equation (19) represents a solution to the inversion problem since it summarizes all information about f. For obtaining an implementable algorithm, however, one attempts to find the estimate $\hat{f}$ that maximizes the *a posteriori* density $p(f/g)$, i.e.

$$\hat{f} = \arg\max_f \ p(f/g) \ .$$

Since $p(g)$ in Equation (19) is independent of f, this reduces to

$$\hat{f} = \arg\max_f \ [p(g/f)p(f)].$$

Thus, depending on the statistical models used for $p(g/f)$ and $p(f)$, different estimates $\hat{f}$ can be obtained by solving this maximization problem. This approach is the framework in which most of the recent developments in image restoration and superresolution are taking place. Solution of the maximization problem with probabilistic models for both $p(g/f)$ and $p(f)$ is referred to as a MAP estimate, while solution with a probabilistic model for $p(g/f)$ while considering the prior as a deterministic quantity leads to a ML estimate. A powerful iterative procedure called "Expectation-Maximization" (EM) algorithm is typically used in formulating specific algorithms.

Two recent algorithms have been receiving a greater degree of attention for their restoration and superresolution capabilities. One of these is the ML algorithm [25]

$$\hat{f}_{k+1}(j) = \hat{f}_k(j)\left[\left\{\frac{g(j)}{\hat{f}_k(j)\otimes h(j)}\right\}\otimes h(j)\right] \quad , \quad j = 1, 2, \ldots, N \qquad (20)$$

where k and $k+1$ denote iteration numbers, $\hat{f}(j)$ denotes the estimate for $f(j)$, the number of photons emitted by the j-th sample of the object (which is a random variable), $g(j)$ denotes the j-th pixel value in the image, $h(j)$ is the j-th element of the sensor PSF and $\otimes$ denotes discrete convolution. This algorithm is developed using Poisson distribution models and has been analytically proven to converge. The second algorithm is the MAP algorithm [14]

$$\hat{f}_{k+1}(j) = \hat{f}_k(j)\exp\left[\left\{\frac{g(j)}{\hat{f}_k(j)\otimes h(j)} - 1\right\}\otimes h(j)\right], \quad j = 1, 2, \ldots, N \qquad (21)$$

which is also developed using Poisson models for both $p(f)$ and $p(g/f)$. The major advantage of a MAP solution is the flexibility in forming the models for the prior $p(f)$, which can greatly help guide the result in a desirable direction in specific applications.

Successful restoration and superresolution using Bayesian methods can greatly depend on the question-- which model to use for the prior? Geman and Geman [26] suggest the use of Markov Random Field (MRF) for constructing prior models which provide the ability to describe spatial correlation in the object intensity function. In a Markov model, each image pixel is envisioned to loosely belong to some connected set of pixels and a MRF prior can be formulated to specify any known information in the reconstruction by exploring the connectivity of pixels. Gibbs functions provide a powerful representation for MRF priors [26].

3.5   Kalman Filtering Methods

When the object and image vectors f and g are considered as stochastic processes and the criterion for image restoration is posed as minimizing the objective function in Equation (8) given as

$$J(g,f) = E\left(\|g - Hf\|^2\right) \qquad (22)$$

where $E(.)$ denotes the expected value, one obtains the minimum mean square estimate (MMSE) $\hat{f}$ upon solving the optimization problem. An efficient scheme to build up the MMSE through a recursive algorithm is given by the Kalman filtering approach. It is based on a state-space representation of the imaging system which can be obtained [27] by augmenting the image formation model given by Equation (4) with an appropriately selected "object model" as a Gauss-Markov random field represented by an autoregressive (AR) process of the form

$$f_i = \sum_j a_j f_{i-j} + u_i \quad , \quad i = 1, 2, \ldots, N \qquad (23)$$

where a_j denotes the AR coefficients and u_i denotes the modeling error. Due to page limitations, details of this approach will not be given here. Some details may however be found in [27-29].

Although early applications of Kalman filtering to the image restoration problem were restricted to one-dimensional formulations, extensive work has been done recently not only in obtaining two-dimensional versions of the filter but also in reducing the computational complexity in their implementation. Two notable results are the Reduced Update Kalman Filter (RUKF), developed by Woods and Radewan [29], where the update procedure is

limited to only those elements of the state vector in a neighborhood of the pixel currently being processed, and the parallel Kalman filtering approach, suggested by Biemond et.al. [30], in which the overall estimation scheme consists of a set of parallel Kalman filters operating on the columns of the image after its rows are decorrelated by a Fourier transform.

4. IMPORTANT CONSIDERATIONS IN TAILORING SUPERRESOLUTION ALGORITHMS

4.1 Requirements for Multispectral Seeker Applications

A number of different approaches to restoration and superresolution of image data were discussed in the last section. Each of these has its own advantages and disadvantages. For an intelligent selection of the right approach in a specific seeker application, it is instructive to consider some basic requirements that need to be met in these applications. These are listed in the following:

1. Flexibility for application to images from different sensing modalities;

2. Performance robustness to tolerate modeling uncertainties, parameter inaccuracies and nonlinearities;

3. Computational requirements that can be met in typical real-time applications;

4. Ensure desired level of resolution enhancement in the presence of significant noise levels;

5. Ensure satisfactory performance in realistic clutter scenarios (with signal-to clutter ratios (SCR) in the range 5-10dB)

6. Facilitate sensor fusion in multi-sensor environments.

The selected approach should permit tailoring an algorithm for efficient processing in the light of these requirements.

The limits on complexity and computational requirements are evident, given the real time operation requirement for a missile seeker. The direct (noniterative) approaches provide obvious advantages over iterative approaches on this count. Nevertheless, the rather poor performance the direct approaches are capable of providing, particularly arising from the sensitivity to noise, clutter and parameter uncertainties, almost always makes them unattractive to use as will become clear from the later discussion.

The requirements arising from the presence of noise and clutter are also evident from the practical environments in which target detection and classification are to be performed. As noted earlier, there are two main sources of noise in these applications, viz. the receiver noise, whose statistics depend on the type of imaging sensor employed and are usually signal dependent, and the quantization noise, which can be realistically modeled by a zero-mean white Gaussian random field that is independent of the image signal. It is well known that deconvolution methods (particularly those that attempt to implement directly an inverse filter) are highly sensitive to noise and require unrealistically high SNR levels (greater than about 60 dB) for satisfactorily processed images. One may note that typical radiometric images (PMMW images, for instance) have SNR levels of about 20 dB (or less), thus highlighting the importance of this requirement. With direct inversion or iterative deconvolution approaches, the noise sensitivity problem needs to be taken care of by some ad hoc modifications, such as modifying the PSF (as in Gleed and Lettington algorithm [19] discussed earlier) or by terminating the number of iterations in the case of iterative deconvolution methods (at the expense of reduced resolution). Bayesian and Kalman filtering approaches offer better ways of accounting for noise (and clutter) in the tailoring of the algorithm.

Perhaps the most significant advantage Bayesian and Kalman filtering approaches offer is the capability for including an appropriately specified "object model" together with the model representing the image formation process in the development of the algorithm. This is obtained by constructing the prior function $p(f)$ in Bayesian approaches or by constructing an AR model of the form given by Equation (23) in the Kalman filtering approach. Note in contrast that the other approaches are based only on the image formation model and hence lack some flexibility in their being adapted to different sensing modalities. As an illustration of this adaptability, it may be observed that while for optical images Poisson probability distributions are generally recognized as good models for describing the photon statistics of the object, these models may not be appropriate for the representation of scattering at millimeter wavelengths [31]. Thus, in employing the Bayesian approach to tailor a superresolution algorithm, one may use Poisson distribution models for the prior $p(f)$ and the likelihood $p(g/f)$ for processing visual images, whereas a different model (such as a Rayleigh distribution model [31,37]) can be employed for processing PMMW images. The recent work of Geman and Geman [26] in suggesting the use of Markov Random Field (MRF) priors that can be represented by Gibbs functions significantly enlarges the capability of Bayesian approaches in this regard. A similar flexibility is offered by the Kalman filtering approach through appropriate selection of AR model coefficients in the process of developing the state-space model of the imaging system.

Yet another advantage of Bayesian and Kalman filtering approaches is the possibility of including the effects of sensor nonlinearity in the models employed. While this is more direct with Bayesian approaches, it will lead to nonlinear state-space models requiring construction of extended Kalman filters with the latter approach. It should be noted in contrast that inverse filtering methods (both direct and iterative deconvolution) require exclusively linear models of the imaging process for executing the needed computations. Furthermore,the exact values of the sensor PSF matrix H are needed as these schemes generally lack robustness to parameter inaccuracies.

It should be emphasized that realization of the advantages cited above for the Bayesian and Kalman filtering approaches in regard to flexibility of modeling different sensor environments require some dedicated effort in the identification of model parameters from experimental data. A considerable literature exists on the identification of dynamic models (ARMA models, state-space models) from input-output data [32] and these techniques can be employed for computing the needed parameters. Through cooperative interactions with sensor designers and image acquisition personnel, and collection of pertinent data under various conditions of altitude, depression angle, scanning angle etc. for each specific sensor under consideration, the needed information for using these schemes can be extracted. The outcome of such modeling and identification studies could be look-up tables listing parameter values to be used under different operating conditions for the various sensors.

The role of superresolution processing in facilitating sensor fusion in multi-sensor environments will be discussed in a later section.

4.2 Optimization of Superresolution Architectures

The importance of tailoring an optimal algorithm that is tuned to the specific sensor characteristics and image acquisition conditions has been emphasized before. A number of factors could be utilized in developing an optimized architecture for the processing of specific image data at hand. One important factor is the *a priori* knowledge regarding the targets of interest or the scene being imaged. This has been adequately discussed in the earlier sections. Certain other considerations, mainly stemming from the discussion on the limits to superresolution

processing given in Section 2.3, can also be used to advantage in realizing desired resolution enhancements with the least computational requirements.

Selection of an adequate sampling rate evidently has a major role in the superresolvability of data. The question of how many samples should be taken across the antenna beamwidth needs to be addressed carefully as it also has an effect on the quantization noise that needs to be accounted for by the superresolution algorithm. Also, as part of the overall processing, upsampling operations (interpolation of data) need to be introduced strategically at various stages of the algorithm since this process expands spatial bandwidth (and supports the increased bandwidth associated with superresolution processing when producing subpixel resolution).

To reduce computational requirements, some forms of preprocessing of data could be employed with the goal of identifying regions of interest (ROI) in the image and requiring the more processing intensive functions of superresolution to be performed only on those regions of the image likely to contain targets of interest. In many applications one will have considerable *a priori* information on the approximate locations or the absence of targets of interest in certain locations which could be utilized in the ROI determination. Appropriately constructed matched filters or neural network-based filters could be employed for a fast execution of this task. Yet another possibility is to use a low-level processing function such as contrast enhancement [33] for sharpening the image before subjecting interesting parts of the image to superresolution processing.

It has been noted in the literature [34] that for sparsely distributed objects on largely zero backgrounds impressive gains in resolution can be achieved from superresolution algorithms. Since many objects of practical interest do not satisfy the constraint of zero background, some intelligent techniques aimed at artificially obtaining these conditions need to be employed. One possibility is to separate the image to be processed into two parts, a background part and a detail part, and decompose the overall estimation problem into estimating the smooth background and then incorporating this knowledge to estimate the "sparsely distributed" detailed portion of the object, as suggested by Frieden and Wells [34]. This has the effect of reducing the spatial extent of the object, and hence benefits superresolution performance as discussed in Section 2.4. A similar idea of background-detail separation in the context of adaptive contrast enhancement of radiological images has been described by Ji et.al. [33].

A number of other ideas and specific techniques aimed at reduction of computational complexity have also been discussed in the image processing literature. While the background-detail separation approach discussed above is one among those that holds particular relevance to superresolution processing in autonomous guidance applications, a few others that can also be employed towards this objective are: (1) controlling the size of pixel neighborhoods (the so-called "cliques") in the application of Bayesian approaches using MRF priors [35], and (2) Reduced Update Kalman Filtering (RUKF) methods where the fact that very few images exhibit large correlation between pixels at large distances (rather, most images show significant correlation only over small neighborhoods of a given pixel) is used to reduce computational complexity. Tactical use of such shortcuts and simplifying approximations can permit tailoring a real-time implementable superresolution processing algorithm in a given application without sacrificing possible performance gains.

4.3   Superresolution Processing in Multisensor Environments and Sensor Fusion Considerations

In a multisensor environment comprising of different sensors operating in different frequency ranges, superresolution processing of image data collected by the various sensors can greatly facilitate sensor fusion goals. Although one of the benefits of employing complementary sensors in a common seeker is to reduce the need for extensive processing of the individual images, attempting to improve the resolution of each image can significantly aid in accomplishing the ultimate goal of a tactical multisensor system, viz. accomplish the mission with an increased probability of survival. It is well known that an efficient approach to realize this goal is to have sensors that can operate in both "autonomous" and "integrated" modes. Such an architecture can provide not only improved fault tolerance and resistance to countermeasures (due to more sensors deployed) but also serves to enhance the target detection and classification capabilities. The system can be designed to be highly responsive by optimizing the capability of each sensor to perform classification and tracking of targets. Then by combining the results of all sensors in the integrated mode operation, a better understanding of the scenario can be developed for the overall surveillance. A logical mechanism for accomplishing these goals is through superresolution processing of each image data by tailoring an appropriate algorithm for each sensor and its operating conditions.

The benefits of superresolution processing in a multisensor environment also come from another aspect. It is well known that the key problem in multisensor fusion is correspondence (or registration) of images and the resulting complexity in selecting and handing over from one sensor to another as the target is approached. For illustration, in a dual mode seeker that combines a IIR with an active MMW radar, a large disparity in the resolution of the two sensors can severely complicate the fusion of the two sensors which may in turn compromise the mission goals.

The availability of different images, typically at different resolution levels, in such dual-mode (or multi-mode) seeker environments presents an unusual opportunity for better quantifying the *a priori* knowledge so critical for the performance of superresolution algorithms. A question of particular interest in this context arises from the possibility of "data fusion", viz., how to utilize the IIR data, for instance, to simplify the processing requirements for MMW images or to enhance the processing performance for these images. Such mechanisms designed to exploit the synergy between the various sensing modalities evidently are capable of offering better overall performance with reduced processing requirements, particularly in missions involving the targeting of mobile missiles and operation in adverse weather environments.

5.   CONCLUSIONS

The principal conclusions and major recommendations arising from this study are briefly summarized in the following:

1.   Superresolution of image data requires bandwidth extrapolation in addition to passband restoration and consequently needs nonlinear processing techniques to be appropriately utilized.

2.   Recreating the frequency components not present in the image and lost due to diffraction limited imaging process is possible and data analysis can indicate existing limits on superresolvability.

3.   Efficient utilization of *a priori* knowledge about the object being imaged is of central importance in our ability to tailor good superresolution algorithms for given data.

4. It is unrealistic to expect any algorithm to work equally well for all types of images. In fact, an optimal algorithm needs to be tailored for each type of sensor and its operating conditions in order to ensure the best resolution improvement performance.

5. Proper tailoring of an image restoration and superresolution algorithm requires some dedicated modeling and identification effort. Any possible interactions with sensor design and development teams are very useful in this step.

6. Optimization of algorithm architecture is necessary for crafting efficient processing algorithms that can be deployed in multispectral seeker environments. A number of tools and concepts exist to assist in this task, utilization of which is critical for realizing attractive operational features in the algorithm designed for each type of sensor and for ensuring its implementational abilities.

REFERENCES

1. B.M.Sundstrom and B.W.Belcher, "Smart tactical autonomous guidance", *Proc. of AIAA Missile Sciences Conference*, Feb. 1992.

2. S.Worrell, "Passive millimeter wave imaging sensor assessment", *GACIAC Publication SR-93-03*, Final Report for Wright Laboratory armament Directorate, Eglin AFB, FL, 1993.

3. W.K.Pratt, "Vector-space formulation of two-dimensional signal processing operations", *Computer Graphics and Image Processing*, Vol.4, pp. 1-24. 1975.

4. H.C.Andrews and B.R.Hunt, *Digital Image Restoration*, Prentice-Hall: Englewood Cliffs, NJ, 1977.

5. W.K.Pratt, *Digital Image Processing*, Wiley: New York, 1978.

6. C.K.Rushforth and J.L.Harris, "Restoration, Resolution and Noise", *J. of Optical Society of America*, Vol. 58, pp. 539-545, 1968.

7. D.O.Walsh and P.A.Delaney, "Direct method for superresolution", *J .of Optical Society of America A*, Vol. 11, pp. 572-579, 1994.

8. R.W.Gerchberg, "Superresolution through error energy reduction", *Optica Acta*, Vol. 21, pp. 709-720, 1974.

9. A.Papoulis, "A new algorithm in spectral analysis and band-limited extrapolation", *IEEE Trans. on Circuits and Systems*, Vol. CAS-22, pp. 735-742, 1975.

10. K. Miller, "Least-squares method for ill-posed problems with a prescribed bound", *SIAM J. Math. Anal.*, Vol. 1. pp. 52-74, 1970.

11. D. Terzopoulis, "Regularization of the inverse visual problem involving discontinuities", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 1, pp. 413-424, 1986.

12. E.L.Kosarev, "Shannon's superresolution limit for signal recovery", *Inverse Problems*, Vol. 6, pp. 55-76, 1990.

13. A.Oppenheim and R.W.Schafer, *Diserete-time Signal Processing*, Prentice-Hall, 1989.

14. P.J.Sementilli, B.R.Hunt and M.S.Nadar, "Analysis of the limit to superresolution in incoherent imaging", *J. of Optical Society of America A,*, Vol. 10, pp. 2265-2276, 1993.

15. E.S.Meinel, "Origins of linear and nonlinear recursive restoration algorithms", *J. of Optical Society of America A*, Vol. 3, pp. 787-799, 1986.

16. R.Prost and R.Goutte, "Discrete constrained iterative deconvolution algorithms with optimized rate of convergence", *Signal Processing*, Vol. 7, pp. 209-230, 1984.

17. M.I.Sezan and A.M.Tekalp, "Survey of recent developments in digital image restoration", *Optical Engineering*, Vol. 29, pp. 393-404, 1990.

18. J.Biemond, R.L.Lagendijk and R.M.Mersereau, "Iterative methods for image deblurring", *Proc. of IEEE*, Vol. 78, pp. 856-883, 1990.

19. D.G.Gleed and A.H.Lettington, "Application of superresolution techniques to passive millimeter-wave images", *Proc. of SPIE Conf. on Applications of Digital Image Processing*, Vol. 1567, pp. 65-72, 1991.

20. R.Bellman, *Matrix Analysis*, McGraw-Hill, 1968.

21. W.H.Press, B.Flannery, S.Teukolsky and W.Vetterling, *Numerical Recipes*, Cambridge Univ. Press: Cambridge, 1986.

22. R.W.Schafer, R.M.Mersereau and M.A.Richards, "Constrained iterative restoration algorithms", *Proc. of IEEE*, Vol. 69, pp. 432-450, 1981.

23. M.A Richards, " Iterative noncoherent angular superresolution", *Proc. of 1988 IEEE National Radar Symp.*, pp. 100-105, April 1988.

24. Z.Ding, "Resolution enhancement of passive millimeter-wave imaging", *Final Report for Summer Faculty Research Program*, AFOSR, August 1993.

25. L.Shepp and Y.Vardi, " Maximum likelihood reconstruction in positron emission tomography" *IEEE Trans. on Medical Imaging*, Vol. 1, pp. 113-122, 1982.

26. S.Geman and D.Geman, " Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 6, pp. 721-741, 1984.

27. N.E.Nahi and C.A.Franco, "Recursive image enhancement by vector scanning", *IEEE Trans. on Communications*, Vol. 21, pp. 305-311, 1973.

28. A.O. Aboutalib, M.S. Murphy and L.M. Silverman, "Digital restoration of images degraded by general motion blurs", *IEEE Trans. Auto. Control*, Vol. AC-22, pp. 294-302, 1977.

29. J.W.Woods and C.A.Radewan, " Kalman filtering in two dimensions", *IEEE Trans. on Information Theory*, Vol. 23, pp. 473-482, 1977.

30. J.Biemond, J.Rieske and J.J.Gerbrands, " A fast Kalman filter for images degraded by both blur and noise", *IEEE Trans. on Acoustics, Speech and Signal Processing*, Vol. 31, pp. 1248-1256, 1983.

31. F.Ulaby, T.F.Haddock and R.T.Austin, " Fluctuation statistics of millimeter-wave scattering from distributed targets", *IEEE Trans. on Geoscience and Remote sensing*, Vol. 26, pp. 268- 281, 1988.

32. L. Ljung and T. Soderstrom, *Theory and Practice of Recursive identification*, MIT Press: Cambridge, 1983.

33. T.L Ji, M.K.Sundareshan and H.Roehrig, " Adaptive image contrast enhancement based on human visual properties", *IEEE Trans. on Medical Imaging*, Vol. 13, pp. 573-587, 1994.

34. B.R.Frieden and D.C.Wells, " Restoring with maximum entropy: Point sources and backgrounds", *J.of Optical Society of America*, Vol. 68, pp. 93-103, 1978.

35. T.J.Hebert and S.S.Gopal, " The GEM MAP algorithm with 3-D SPECT system response", *IEEE Trans. on Medical Imaging*, Vol. 11, pp. 81-90, 1992.

36. A.B.Mahmoodi and M.Kaveh, " Signal processing considerations for a millimeter-wave seeker", *Proc. of SPIE 423*, (Ed. J.T. Wiltse), 1983.

ANALYZING CONSTANT FALSE-ALARM RATE
SAR IMAGE TARGET DETECTORS

John A. Tague
Associate Professor
School of Electrical Engineering and Computer Science

Ohio University
Athens, Ohio

ANALYZING CONSTANT FALSE-ALARM RATE
SAR IMAGE TARGET DETECTORS

John A. Tague
Associate Professor
School of Electrical Engineering and Computer Science
Ohio University

Abstract

Many automatic target recognition systems use a prescreening algorithm to identify image segments which may contain interesting objects. Its performance can be quantified by a receiver operating characteristic - a plot of detection probability versus number of false alarms per square kilometer of imagery. Recently, a constant false alarm rate prescreener was tested on images formed at several different resolutions. We can be confident that the experimental results are not an "artifact" of the particular data set. Furthermore, under certain conditions, theory validates the experimental trends.

ANALYZING CONSTANT FALSE-ALARM RATE
SAR IMAGE TARGET DETECTORS

John A. Tague

1. Introduction

Automatic target recognition (ATR) systems which process synthetic aperture radar (SAR) images can be implemented by a pipelined set of processors [1]. The first processing stage, called detection or prescreening, searches through the image and identifies picture elements (pixels) which may belong to a target of interest. These "target-like" pixels are grouped together by a clustering algorithm, and the resulting "regions of interest" are passed to the next stage of processing. The goal of the prescreener is to identify all actual targets located in the scene and reject all false targets - collections of bright spots which appear target-like but are actually "clutter." Clutter is produced by natural and man-made objects which are imaged by the radar but are not targets of interest. Within the context of the target recognition problem, it confounds our ability to locate targets of interest within a SAR image and classify them correctly. Distinguishing bright spots produced by targets from bright spots produced by clutter is a challenging problem for the ATR system designer.

This report studies the prescreening process and how it is affected by the resolution of the images under consideration. The prescreener is part of a larger system and, as such, it is important for the system designer to understand how its performance is affected by changing design parameters in other parts of the system. A fundamental design parameter which impact the entire system is image resolution. Generating high resolution images is a costly undertaking; therefore, it is of great practical concern to quantify how resolution affects detection performance.

Recently, a constant false alarm rate (CFAR) prescreener and a target classifier were tested experimentally on a large set of field data [2]. The specific goal of the study was to examine how resolution impacts the performance of both of these algorithms. The classifier performance results were as expected: high resolution images yielded the best classifier performance. On the other hand, the prescreener results were surprising: its performance did not improve as the image resolution was varied from coarse to fine. This study, which is motivated by the preceding work, aims at answering two questions. First, can the experimental results be attributed to chance; that is to say, are the results merely an "artifact" of the particular data set? Second, can the results be predicted from a theoretical analysis of the problem? If we can answer these questions to our satisfaction, we can have confidence that the experiment was carried out properly.

This report is organized as follows. In the next section, we describe the image under examination and the prescreening algorithm. In Section 3, we formulate the measurement models which form the basis for the subsequent performance analysis. Next, in Section 4, we show how to compute detection and false alarm probabilities. We find closed-form solutions for both and show how they can be evaluated numerically. In Section 5, we examine resolution spoiling and how it affects the target and clutter pixel amplitudes. An example is presented in Section 6. Section 7 is devoted to confidence interval estimation, and we present our conclusions in Section 8.

2. The Prescreening Algorithm

The image which is passed to the prescreener is a synthetic aperture radar image of a region several square kilometers in area. In their raw form, the image pixels are complex-valued, and in principle, they could be processed by the system. However, experience has shown that better results are obtained by processing "quarter-power" images. The quarter-power image is generated by taking the square root of the magnitude of each complex pixel. We will describe the motivation behind this transformation in greater detail later in the report.

The prescreener examines pixels one at a time. Suppose that p_{ij} is the pixel under examination. We compare it to its neighbors by way of the following scheme. Surrounding the test pixel is a guard region, a square whose sides are twice that of potential targets. Pixels within the guard region are not used because they may be target-like. Pixels within in a frame on the guard region boundary are used to estimate the local clutter mean and standard deviation. The frame area is kept constant, meaning that the number of pixels depends on the resolution. A total of N_{cells} is used to estimate the local clutter mean and standard deviation.

The test statistic, which measures how much the test pixel stands out above its neighbors, is defined as follows:

$$z = \frac{p_{ij} - \widehat{\mu}_Q}{\widehat{\sigma}_Q}. \tag{1}$$

In Equation (1), $\widehat{\mu}_Q$ and $\widehat{\sigma}_Q$ denote the sample mean and sample standard deviation of the pixels in the boundary region. We declare that the pixel is "target-like" if $z \geq \gamma$; on the other hand, if $z < \gamma$ we decide that the pixel is "clutter-like." All pixels declared target-like are identified as such, and a clustering algorithm is used to group them into chips which are called "regions of interest."

The threshold γ determines the false alarm rate of the prescreener. It can be proven that for a given threshold, the prescreener's false alarm rate is independent of the clutter mean and variance when the clutter pixels are independent, identically distributed, Gaussian random variables. Such a processor is said be a constant false alarm rate (CFAR) detector.

3. Problem Formulation

We call on the prescreener to decide if the pixel under examination is "clutter-like" or "target-like." Random variations in clutter and target pixel amplitudes introduce uncertainty into the decision-making process. Therefore, we will pose the detection problem within the framework of statistical decision theory.

We define the null hypothesis, denoted H_0, to mean that the pixel under examination is clutter-like. Under H_0,

$$p_{ij} = n_{ij}, \tag{2}$$

where n_{ij} is a Gaussian distributed random variable with mean μ_Q and variance σ_Q^2.

The statistical properties of the quarter-power clutter can be obtained analytically as follows. Let us assume that complex clutter pixels are independent, identically distributed complex-valued Gaussian random variables with zero mean and variance σ^2. Given this model for the complex clutter, the quarter-power clutter probability density function is easily obtained. It is very close to Gaussian with mean $\mu_Q \approx 0.9064\sqrt{\sigma}$ and variance $\sigma_Q^2 \approx 0.06463\sigma$. This property motivates its use in target recognition problems. Furthermore,

assuming the complex clutter pixels are independent and identically distributed, so are the quarter-power clutter pixels.

We define the alternate hypothesis H_1 to mean that the pixel is target-like. Under H_1,

$$p_{ij} = \sqrt{s_{ij}} + n_{ij}, \tag{3}$$

where $n_{ij} \sim N(\mu_Q, \sigma_Q^2)$ and $\sqrt{s_{ij}}$ is the deterministic target return. The n_{ij} term models random fluctuations in the return.

4. Performance Analysis

Prescreening performance is usually quantified by a receiver operating characteristic – a plot of detection probability versus number of false alarms per square kilometer of processed imagery. In this section, we carry out the calculations which lead to closed-form solutions for each of these performance criteria.

The detection probability, denoted P_d, is the probability that we decide H_1 when H_1 is actually correct. It is computed from the formula

$$P_d = \int_\gamma^\infty f(z|H_1)\, dz, \tag{4}$$

where $f(z|H_1)$ is the probability density function of z given H_1. The false alarm probability, denoted P_{fa}, is the probability that we decide H_1 when H_0 is actually correct. It is computed from the formula

$$P_{fa} = \int_\gamma^\infty f(z|H_0)\, dz, \tag{5}$$

where $f(z|H_0)$ is the probability density function of z given H_0.

The heart of the performance analysis problem is to find $f(z|H_0)$ and $f(z|H_1)$ given the test statistic and measurement models set forth in Sections 2 and 3. Let us begin with the false alarm probability calculation. Recall that N_{cells} are used to estimate μ_Q and σ_Q. It is easy to show that

$$\frac{p_{ij} - \widehat{\mu}_Q}{\sigma_Q} \sim N(0,\ (N_{cells} + 1)/N_{cells}) \tag{6}$$

and that

$$\widehat{\sigma}_Q^2/\sigma_Q^2 \sim \chi^2_{N_{cells}-1}. \tag{7}$$

Now z is almost t–distributed. If we scale z and define

$$w = \sqrt{\frac{N_{cells}}{N_{cells} + 1}}\, z, \tag{8}$$

it can be proven that w has a central t distribution with $N_{cells} - 1$ degrees of freedom [3] [4]. Therefore,

$$P_{fa} = P(z \geq \gamma \mid H_0) = 1 - F_w(\sqrt{N_{cells}/N_{cells} + 1}\ \gamma|H_0), \tag{9}$$

where $F_w(\cdot \mid H_0)$ is the cumulative distribution function of w. If we need the number of false alarms per square kilometer, we simply multiply P_{fa} by the number of pixels per square kilometer of image.

The detection probability calculation proceeds along similar lines. As before, N_{cells} are used to estimate μ_Q and σ_Q. It is easy to show that

$$\frac{p_{ij} - \widehat{\mu}_Q}{\sigma_Q} \sim N(\sqrt{s_{ij}/\sigma_Q^2},\ (N_{cells} + 1)/N_{cells}) \tag{10}$$

and that

$$\hat{\sigma}_Q^2 / \sigma_Q^2 \sim \chi_{N_{cells}-1}^2. \tag{11}$$

The probability density function of z is nearly a non-central t distribution [4]. If we scale z and define

$$w = \sqrt{\frac{N_{cells}}{N_{cells}+1}} \; w \tag{12}$$

then w has a non-central t distribution with $N_{cells} - 1$ degrees of freedom and non-centrality parameter

$$\delta = \sqrt{\frac{N_{cells}}{N_{cells}+1}} \; \sqrt{\frac{s_{ij}}{\sigma_Q^2}}. \tag{13}$$

Therefore,

$$P_d = P(z \geq \gamma \mid H_1) = 1 - F_w(\sqrt{N_{cells}/N_{cells}+1} \; \gamma | H_1), \tag{14}$$

where $F_w(\cdot \mid H_1)$ is the cumulative probability distribution function of w.

The functional forms of $F_w(\cdot \mid H_0)$ and $F_w(\cdot \mid H_1)$ are very complicated, and many different techniques have been proposed to evaluate them numerically [5] [6]. The following "recipe" appears to be the easiest to implement. First, if the degrees of freedom $f = N_{cells} - 1$ is odd, then

$$F_w(t) = Pr(w \leq t) = G(-\delta\sqrt{B}) + 2T(\delta\sqrt{B}, A) + 2(M_1 + M_3 + \ldots + M_{f-2}), \tag{15}$$

where δ is the non-centrality parameter,

$$A = \frac{t}{\sqrt{f}} \qquad \text{and} \qquad B = \frac{f}{f+t^2}, \tag{16}$$

$$G'(x) = \frac{1}{2\pi} e^{-x^2/2}, \tag{17}$$

$$G(x) = \int_{-\infty}^{x} G'(t) \; dt, \tag{18}$$

and

$$T(h, a) = \frac{1}{2\pi} \int_0^a \frac{\exp(-h^2(1+x^2)/2)}{1+x^2} \; dx. \tag{19}$$

This integral must be evaluated numerically. The M's are defined below. For even values of f,

$$F_w(t) = Pr(w \leq t) = G(-\delta) + \sqrt{2\pi}(M_0 + M_2 + \ldots + M_{f-2}), \tag{20}$$

where

$$M_{-1} = 0, \tag{21}$$

$$M_0 = A\sqrt{B} \; G'(\delta\sqrt{B}) \; G(\delta A\sqrt{B}), \tag{22}$$

$$M_1 = B \left(\delta A M_0 + \frac{A}{\sqrt{2\pi}} G'(\delta) \right), \tag{23}$$

$$M_2 = \frac{1}{2} B \left(\delta A M_1 + M_0 \right), \tag{24}$$

$$M_3 = \frac{2}{3} B (\delta A M_2 + M_1); \tag{25}$$

and in general, for $k \geq 4$,

$$M_k = \frac{k-1}{k} B (a_k \, \delta \, A M_{k-1} + M_{k-2}), \tag{26}$$

where

$$a_k = \frac{1}{(k-2)a_{k-1}} \quad \text{for} \quad k \geq 3 \quad \text{and} \quad a_2 = 1. \tag{27}$$

The preceding equations can be used to calculate P_{fa} as well as P_d. The false alarm rate requires evaluating the distribution of a central t distributed random variable; this is obtained by setting the non-centrality parameter $\delta = 0$.

We point out that if N_{cells} is large, say > 100, then we can approximate the central and non-central t distributions by normal distributions [5]. Under these conditions,

$$P_d \approx 1 - \Phi(\gamma - \delta) \tag{28}$$

and

$$P_{fa} \approx 1 - \Phi(\gamma). \tag{29}$$

We are now equipped to calculate detection and false alarm probabilities as a function of N_{cells}, clutter power, and target amplitude. However, before we can compare detection performance as a function of image resolution, we need to understand how resolution affects each of these parameters. We take this up in the next section.

5. Resolution Spoiling and Signal-to-Noise Ratios

Let us examine how resolution affects the signal-to-noise ratio, a critical performance parameter. To do so, we must begin by describing the signal processing procedure which was used to alter image resolution throughout the experiment.

High resolution SAR image data was provided to the investigators, and signal processing was used to reduce the cell resolution size when required. This process, called resolution spoiling, was implemented as follows. First, a two-dimensional fast Fourier transform of the complex image was computed. Next, those DFT bins corresponding to high spatial frequencies were discarded according to specifications. Finally, an inverse DFT of the low spatial frequency bins was computed, producing a low resolution image. Both images were critically sampled and, as such, had the same scene sizes. Only the resolution cell size was changed.

How does resolution spoiling affect SAR images? Consider the one-dimensional image of a point scatterer located at $x = 0$:

$$s(x) = \frac{A \sin(2\pi B x)}{\pi x}. \tag{30}$$

The independent variable x is cross range in meters, $A > 0$ is a scaling factor, and the spatial bandwidth of $s(x)$ is $-B \leq k_x \leq +B \ m^{-1}$. The maximum value of $s(x)$ is $s(0) = 2AB$. Now suppose the image resolution is cut by a factor of two; that is to say, its bandwidth is reduced to $-B/2 \leq k_x \leq +B/2 \ m^{-1}$. Then the peak amplitude falls by a factor of two: $s(0) = AB$. If this peak were centered in the test cell, we may conclude that in general, resolution spoiling reduces image amplitude by a factor equal to the spoiling factor.

But is this always the case? Suppose that we process a one-dimensional image of two point scatterers located at $x = 0$ and $x = 1$. Then

$$s(x) = \frac{A\sin(2\pi Bx)}{\pi x} + \frac{A\sin(2\pi B(x-1))}{\pi(x-1)}. \tag{31}$$

Let $B = 0.5 \ m^{-1}$ and $A = 1$. We sample $s(x)$ at the Nyquist interval which is every meter; this yields the sampled image as illustrated in Figure 1. The resolution cell size is 1 meter.

We now proceed with the DFT processing required to increase the resolution cell size to 2 meters. In our numerical experiment, we computed a 512-point FFT of the high resolution sampled image, pulled out the 256 bins corresponding to high spatial frequencies, and then took a 256-point inverse FFT in order to compute the spoiled image. The result is shown in Figure 2. If we compare it to the actual low resolution image, illustrated in Figure 3, we find that the samples are scaled by an incorrect factor of two. Why? We forgot to account for the fact that we are processing analog domain waveforms using discrete-time signal processing. We must divide the inverse FFT output by two in order to obtain correct analog domain amplitudes. Figure 4 shows the correctly scaled answer; the sample values interpolate the low resolution image correctly. Also notice that $s(0)$ in Figure 4 is approximately 0.8, not too much less than $s(0)$ in the high resolution image, which is 1.0. If $s_{ij} = s(0)$ in the low and high resoluton images, resolution spoiling does not reduce s_{ij} by the factor we may expect.

How does resolution spoiling affect clutter power? Suppose we are given a high resolution, one-dimensional complex clutter image containing N pixels. Each pixel has variance σ^2. We spoil cell resolution by a factor of two using the same signal processing steps described above. Let x be an $N \times 1$ vector containing the complex clutter pixels. Define I_M as the $M \times M$ identity matrix, and let $0_{M \times N}$ be an $M \times N$ matrix of all zeros. The DFT operation can be represented as the matrix-vector product

$$X = W_N x, \tag{32}$$

where $X =$ the $N \times 1$ vector of DFT coefficients and W_N is the $N \times N$ DFT matrix.

The spoiling operation can be represented in terms of the matrix-vector product

$$Y = SX = SW_N x, \tag{33}$$

where

$$S = \begin{bmatrix} I_{N/4 \times N/4} & 0_{N/4 \times N/2} & 0_{N/4 \times N/4} \\ 0_{N/4 \times N/4} & 0_{N/4 \times N/2} & I_{N/4 \times N/4} \end{bmatrix} \tag{34}$$

is the $N/2 \times N$ resolution spoiling matrix. The spoiled image is obtained by computing the $N/2$ point inverse DFT of Y and scaling the result:

$$y = (1/2)W_{N/2}^{-1}Y = (1/2)(2/N)W'_{N/2}Y = (1/N)W'_{N/2}SW_N x, \tag{35}$$

where prime denotes Hermitian transpose and the $1/2$ scales the spoiled image amplitudes correctly.

It is easy to show that the spoiled clutter pixels are zero mean and Gaussian distributed. The covariance matrix of y can be computed with a bit of extra effort:

$$R_y = E\{yy'\} = (1/N^2)E\{W'_{N/2}SW_N xx'W'_N S'(W'_{N/2})'\}$$

$$(1/N^2)W'_{N/2}SW_N E\{xx'\}W'_N S'W_{N/2}\}$$

$$= (\sigma^2/N)W'_{N/2}SS'W_{N/2}. \tag{36}$$

We obtained this result because $W'_N W_N = NI_N$ and $E\{xx'\} = \sigma^2 I_N$. Next, it is easy to show that $SS' = I_{N/2 \times N/2}$. Therefore,

$$R_y = (\sigma^2/N)W'_{N/2}W_{N/2} = (\sigma^2/N) \times (N/2)I_{N/2} = (\sigma^2/2)\, I_{N/2}. \tag{37}$$

Reducing cell resolution by a factor of two reduces complex clutter variance by a factor of two.

We now present two examples which illustrate how resolution spoiling affects signal-to-noise ratio. The quarter-power image signal-to-noise ratio (SNR) is

$$\mathrm{SNR} = \frac{s_{ij}}{\sigma_Q^2} = \frac{s_{ij}}{0.06463\sigma}. \tag{38}$$

Suppose that we image a single point scatterer at one meter resolution. Its peak value $s(0) = s_{ij} = 1$ and we set $\sigma^2 = 1$. When the peak lies in the cell under test,

$$\mathrm{SNR} = 10\log_{10}\left(\frac{s_{ij}}{0.06463\sigma}\right) = 10\log_{10}(1/0.06463) = 11.89 \text{ dB}. \tag{39}$$

If we image the same point scatterer at two meter resolution, $s(0) = s_{ij} = 0.5$, and the clutter power drops to 0.5. Therefore,

$$\mathrm{SNR} = 10\log_{10}\left(\frac{0.5}{0.06463 \times 0.7071}\right) = 10.4 \text{ dB}. \tag{40}$$

The signal-to-noise ratio is reduced by approximately 1.5 dB.

Finally, suppose we image two point scatterers, located at $x = 0$ and $x = 1$ meter, at one and two meter resolution. We use $s(0) = s_{ij}$ to compute SNR. As before, we set $\sigma^2 = 1$ at high resolution, and at two meter resolution, the complex clutter power is 0.5. In the 1 meter resolution image,

$$\mathrm{SNR} = 10\log_{10}(1.0/0.06463) = 11.89 \text{ dB}, \tag{41}$$

and in the two meter resolution image,

$$\mathrm{SNR} = 10\log_{10}\left(\frac{0.8}{0.06463 \times 0.7071}\right) = 12.54 \text{ dB}, \tag{42}$$

which is slightly more than the high resolution SNR. This suggests that under certain circumstances, the low resolution system may work slightly *better* than the high resolution system. In the next section, we will show this is indeed the case.

6. Examples

We have written computer programs in the Matlab programming language which compute detection and false alarm probabilities under a wide range of operating conditions. In this section, we will show that under certain conditions, our theory produces results which are consistent with experimental results.

The receiver operating characteristics plot detection probability versus number of false alarms per square kilometer. To calculate the number of false alarms per square kilometer, we assumed that 4.8 square kilometers of clutter imagery were tested. Two resolutions were used: high resolution pixels were 3 meter

by 1 meter and low resolution pixels were 3 meter by 3 meter. In the high resolution examples, we set $N_{cells} = 132$; in the low resolution examples; we set $N_{cells} = 44$.

In the first test, given H_1, the test cell contains a peak corresponding to the image of an isolated point scatterer. Its peak value in the high resolution image is $s(0) = s_{ij} = 1$. In the low resolution image, we reduced it value by a factor of three, so that $s(0) = s_{ij} = 0.3334$. The complex clutter power in the high resolution imagery was $\sigma^2 = 1$; in the low resolution imagery, it was $\sigma^2 = 0.3334$.

Figure 5 displays the results. The solid line is the ROC of the high resolution data and the dashed line is the ROC of the low resolution data. The results are not surprising: the high resolution system outperforms the low resolution system. The signal-to-noise ratio in the high resolution data is 11.89 dB and in the low resolution data it is 9.51 dB. It is interesting to observe that the detector performance is very sensitive to small changes in the signal-to-noise ratio.

In the second test, given H_1, the test cell interpolates the image of two point scatterers located at $x = 0$ and $x = 1$ meters. We computed the images at 1 and 3 meter resolutions and set $s_{ij} = s(0)$. In the high resolution image, $s(0) = s_{ij} = 1$; in the low resolution image, $s(0) = s_{ij} = 0.6$. The complex clutter parameters are identical to the first example.

Figure 6 displays the results. They are interesting, somewhat surprising, but consistent with experimental results: the *low resolution* system outperforms the high resolution system. We conjecture this is tied to SNR, because at high resolution, SNR = 11.89 dB, and at low resolution, SNR = 12.06 db. We have found that the detector performance is sensitive to small changes in SNR and less sensitive to changes in N_{cells}.

How does N_{cells} affect detector performance? We know that more statistically independent clutter pixels are used to estimate μ_Q and σ_Q at high resolutions. This in and of itself improves performance. However, in this particular situation, the clutter variance increases with resolution. We conjecture that the increasing clutter variance "cancels out" the variance reduction attained by averaging across more independent cells at higher resolutions.

7. Confidence Intervals and Experiment Validation

So far, we have sought a model which predicts the performance trends made manifest in the experimental results. However, we have yet to address a more fundamental issue: Can we trust the experimental results? To answer this question, we will describe the "confidence interval" concept, show how they can be computed, and present some interesting results.

The receiver operating characteristics developed in the resolution study are plots of *estimated* P_d and number of false alarms per square kilometer. They come from testing the CFAR processor on thousands of independent pixels, keeping track of "successful" results, and dividing the number of success by number of independent trials. For example, one can estimate P_d at a given detection threshold γ by processing target-like pixels, counting the total number of detections, and dividing by the total number of processed pixels. In addition, we can estimate false alarm probabilities from clutter data using the same technique. We denote the detection and false alarm probability estimates by $\widehat{P}_d$ and $\widehat{P}_{fa}$ respectively.

How do we assess the accuracy of $\widehat{P}_d$ and $\widehat{P}_{fa}$? One approach is to calculate confidence intervals derived from statistical estimation theory [7]. A confidence interval is a region, derived from the experimental result, which contains the true value with probability very close to one. For example, the "99% confidence interval for P_d" is the random interval $(\widehat{P}_d - a, \widehat{P}_d + b)$ such that

$$Pr(\widehat{P}_d - a \leq P_d \leq \widehat{P}_d + b) = 0.99. \tag{43}$$

The 99% confidence interval for P_{fa} is defined the same way.

Let us sketch out the confidence interval calculations for P_{fa}; the derivation for P_d is the same. Suppose that we test N clutter pixels. We declare a false alarm when $z \geq \gamma$. Let X_i = indicator function for ith pixel:

$$X_i = \begin{cases} 1, & Z \geq \gamma; \\ 0, & Z < \gamma. \end{cases} \tag{44}$$

The $\{X_i\}$'s are independent, identically distributed Bernoulli random variables, and

$$\widehat{P}_{fa} = \frac{1}{N} \sum_{i=1}^{N} X_i. \tag{45}$$

Now in this study N is large, and it can be proven that

$$\sqrt{N} \frac{\widehat{P}_{fa} - P_{fa}}{\sqrt{P_{fa}(1 - P_{fa})}} \sim N(0, 1). \tag{46}$$

This mean that

$$Pr\left(\left|\frac{\widehat{P}_{fa} - P_{fa}}{\sqrt{P_{fa}(1 - P_{fa})}}\right| \leq z(1 - 0.5\alpha)\right) = 1 - \alpha. \tag{47}$$

The expression $z(1 - 0.5\alpha)$ is that value of z such that $\Phi(z) = 1 - 0.5\alpha$. With a bit of work, Equation 47 can be reformulated such that a and b emerge from it:

$$a = (S + k_\alpha^2/2 - k_\alpha \sqrt{S(N - S)/N + k_\alpha^2/4})/(N + k_\alpha^2), \tag{48}$$

and

$$b = (S + k_\alpha^2/2 + k_\alpha \sqrt{S(N - S)/N + k_\alpha^2/4})/(N + k_\alpha^2). \tag{49}$$

In Equations 48 and 49, $S = N\widehat{P}_{fa}$ and k_α = that value of z such that $\Phi(z) = 1 - 0.5\alpha$. To find the 99% confidence intervals, we set $\alpha = 0.01$. Then we determine k_α = the value of z such that $\Phi(z) = 1 - 0.005 = 0.995$. From tables of the cumulative distribution function of a unit normal random variable, $k_\alpha = 2.575$.

We will need confidence intervals on the false alarms per square kilometer estimates. They are found by multiplying a and b shown above by the number of pixels per square kilometer at a given cell resolution.

We have written computer programs in the Matlab programming language which calculate the confidence intervals. In the examples which follow, we present confidence intervals for $\widehat{P}_{fa}$ and estimated false alarms per square kilometer. We computed confidence intervals for high and low resolution images: "High resolution" meaning 0.3 by 0.3 meter pixels, and "Low resolution" meaning 3 by 3 meter pixels.

Figures 7 and 8 are the $\widehat{P}_{fa}$ confidence interval plots. The reader should interpret these plots as follows. Consider the low resolution results, and suppose that the outcome of a test yields $\widehat{P}_{fa} = 1 \times 10^{-3}$. Then the probability that P_{fa} is between the interval delineated by the dashed lines is 0.99. We can see that the resolution study data set is large enough to yield good estimates of P_{fa}, even when it is very small.

Figures 9 shows the confidence intervals on the false alarms per square kilometer estimates. Only the high resolution results are shown because they are virtually identical to the low resolution intervals. We have examined the numbers and found that the intervals are actually somewhat tighter on the high resolution data. But pratically speaking, the differences are insignificant.

We conclude from these typical results that enough data was used in the resolution study to provide the investigators with meaningful receiver operating characteristics.
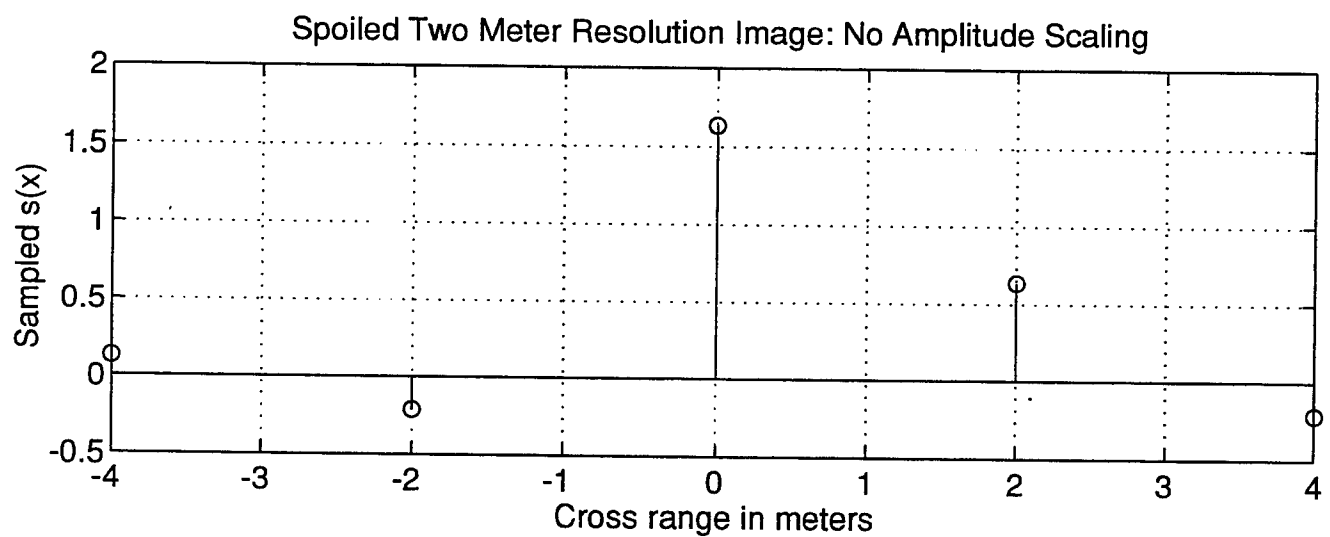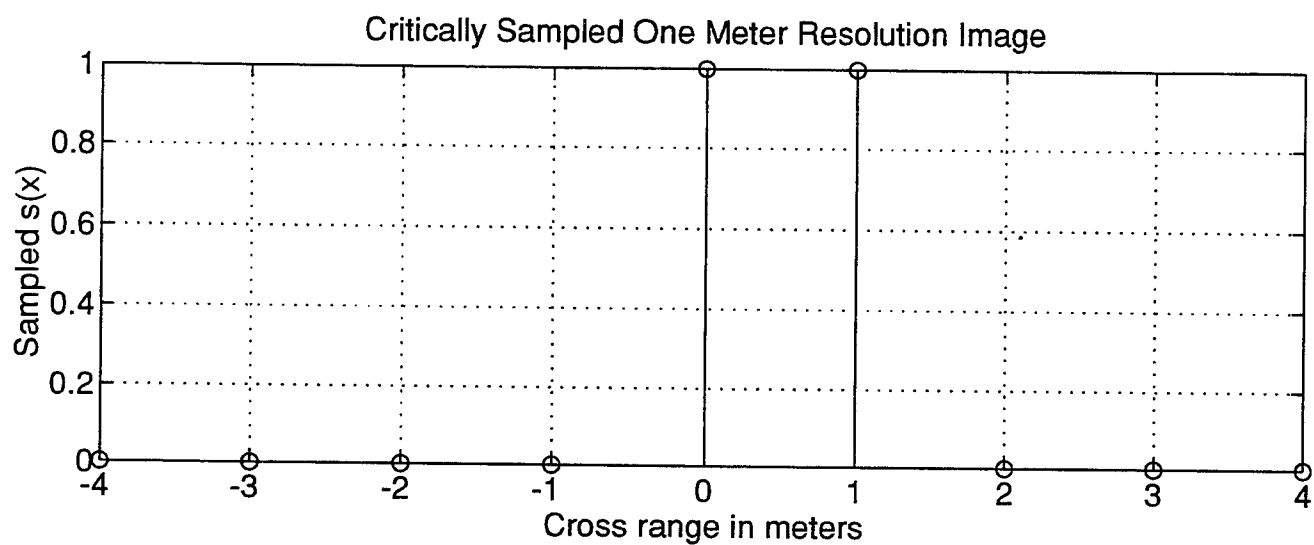
8. Conclusions

We find that a simple data model predicts the trends of the resolution study results. We also conclude that the experimental results are not produced by random happenstance; on the contrary, enough data was used to draw meaningful conclusions regarding the receiver operating characterstics of the CFAR detector.

Although the theory developed in this report predicts the trends of the experimental results, significant discrepencies between the analytical and experimental results remain. One reason for this is related to our model for s_{ij}. We assumed it was deterministic and the same for all target-like cells. Random variations in the amplitude of p_{ij} were modeled by n_{ij}, a random variable whose statistical properties are identical under both hypotheses. We conjecture this model is inadequate, and that s_{ij} itself needs to be modeled as a random variable.

In addition, we believe that a portion of the difference - exactly how much is not known - is due to clutter non-Gaussianity. It would be worthwhile to find a better clutter model, one which explains the clutter properly, and one which should yield analytical results closer to the experimental results.

References

[1] Daniel E. Kreithen *et. al.*, "Discriminating targets from clutter," *Lincoln Laboratory Journal*, Volume 6, Number 1, pp. 25-52, Spring 1993.

[2] Ronald L. Dilsavor *et. al.*, "Mobile target cueing performance at various SAR image resolutions," 41st Tri-Service Symposium, 1995.

[3] Normal I. Johnson and Samuel Kotz, *Distibutions in Statistics: Continuous Univariate Distributions, Part 2*, Houghton Mifflin Company, Chapters 27 and 31, 1970.

[4] Normal I. Johnson and Samuel Kotz, *Distibutions in Statistics: Continuous Univariate Distributions, Part 1*, Houghton Mifflin Company, Chapter 13, 1970.

[5] D.B. Owen, "A survey of properties and applications of the non-central *t*-distribution," *Technometrics*, Volume 10, pp. 445-478, August 1968.

[6] Don B. Owen, "The power of the Student's *t*-test," *Journal of the American Statistical Association*, Volume 60, pp. 320-333, March 1965.

[7] Peter J. Bickel and Kjell A. Doksum, *Mathematical Statistics: Basic Ideas and Selected Topics*, Prentice-Hall, Chapter 5, 1977.

Critically Sampled One Meter Resolution Image

Spoiled Two Meter Resolution Image: No Amplitude Scaling

SAR Image of Two Point Scatterers: Two Meter Resolution
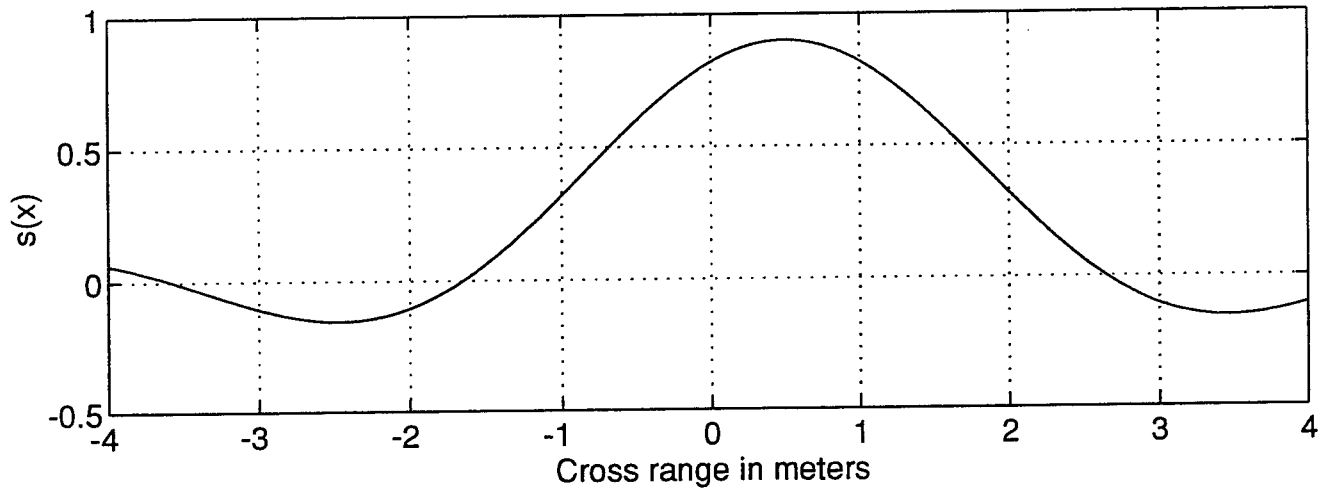
Spoiled Two Meter Resolution Image: Proper Scaling
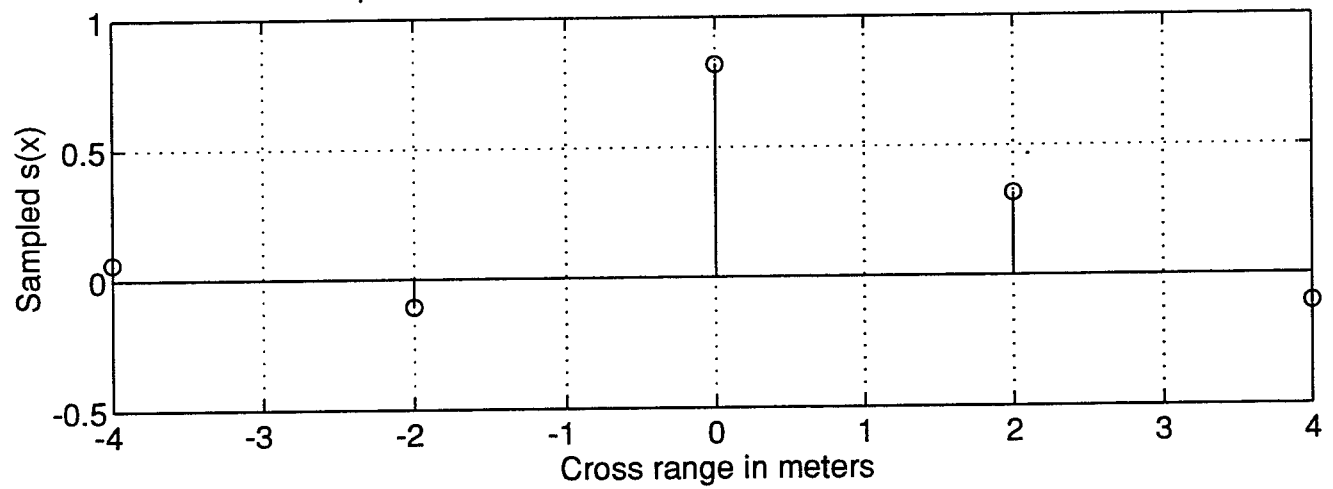
Rural Clutter: One Point Scatterer
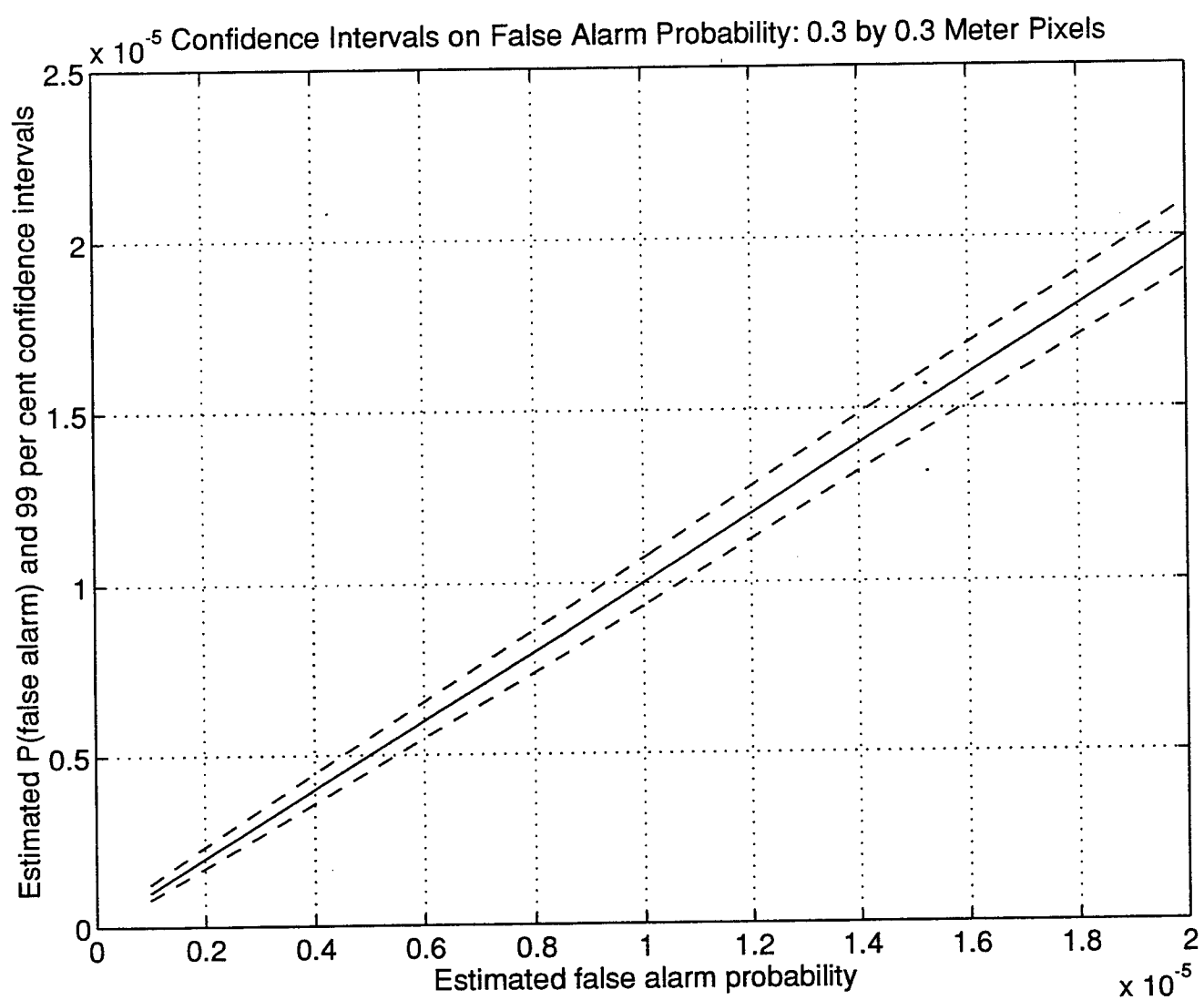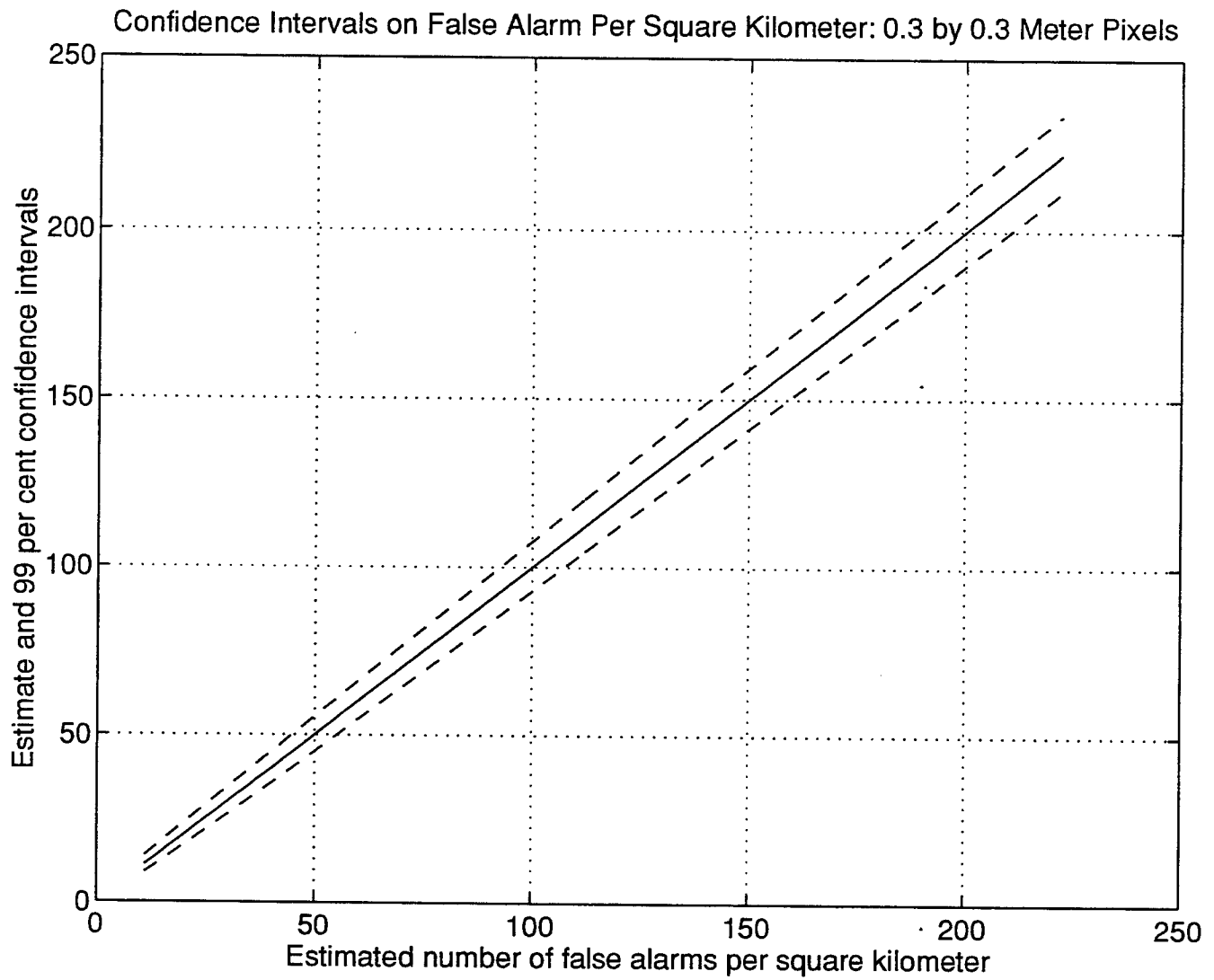
Rural Clutter: Two Point Scatterers

Probability of Detection

Number of false alarms per square kilometer

Confidence Intervals on False Alarm Probability: 3 by 3 Meter Pixels

Confidence Intervals on False Alarm Probability: 0.3 by 0.3 Meter Pixels

Confidence Intervals on False Alarm Per Square Kilometer: 0.3 by 0.3 Meter Pixels

HARD TiC COATINGS PREPARED BY PULSED LASER DEPOSITION
AND
A COMPARISON WITH MAGNETRON SPUTTERED TiC COATINGS

Jinke Tang
Associate Professor
Department of Physics

University of New Orleans
Lakefront
New Orleans, LA 70148

HARD TiC COATINGS PREPARED BY PULSED LASER DEPOSITION
AND
A COMPARISON WITH MAGNETRON SPUTTERED TiC COATINGS

Jinke Tang
Associate Professor
Department of Physics
University of New Orleans

Abstract

Tribological properties of TiC coatings grown by pulsed laser deposition (PLD) and magnetron sputtering were investigated. The PLD TiC coatings grown at room temperature were found to be much harder than the TiC films grown by magnetron sputtering under similar conditions. The hardness of PLD TiC coatings measured by a nanoindentation test was as high as ~40 GPa in contrast to ~20 GPa of the magnetron sputtered ones. The coefficient of friction of the PLD films measured with a pin-on-disk type tribometer had a typical value of about 0.2 when using a 440C stainless steel pin. Scratch tests indicated that magnetron sputtered TiC coatings adhere better to the stainless steel substrate than the PLD grown coatings. The relatively poor adhesion of the PLD coatings obtained from this scratch test was probably partly due to its high hardness. The adhesion of magnetron sputtered TiC coatings could be modified by inserting a metallic interlayer between the coating and stainless steel substrate. Mo interlayer had a detrimental effect on the adhesion caused probably by the poor stress bearing capability of the porous Mo film deposited at low temperature. However, the insertion of both Ti and Cr interlayers enhanced the adhesion of TiC, by as much as 25%.

HARD TiC COATINGS PREPARED BY PULSED LASER DEPOSITION
AND
A COMPARISON WITH MAGNETRON SPUTTERED TiC COATINGS

Jinke Tang

Introduction

Coatings are being used more and more frequently to solve critical tribological problems occurring either in the design phase or during the use of products and production machinery. The areas of application are those where a combination of the properties of both the substrates and coatings is required. Examples are the wear resistance, frictional behavior or tendency for adhesion of coatings, combined with the toughness, machinability, cost or physical properties of the substrate materials.

The United States Air Force is very much interested in the research in the area of coating tribology since wear, corrosion, hardness, friction and adhesion of coatings are the most important factors that determine the performance of aerospace systems operating at high and low temperatures. Such research is especially needed in the design of high temperature engines proposed for the aircraft of the twenty-first century. The Air Force is also interested in coatings that apply to electronic technology and lubrication of spaceborn precision direction mechanisms.

One potential such coating is TiC. TiC is extremely hard and has very low coefficient of friction and high resistance to oxidation [1]. It is semimetallic ceramic possessing good thermal stability and electrical and thermal conductivity [2,3]. Therefore, TiC has been used in a number of applications. It is used as a protective coating to increase the wear life of steel parts under extreme chemical, thermal and mechanical conditions. It is also being studied as a low friction thermal barrier coating for cylinder walls in the adiabatic diesel engines. It has been identified as a candidate material for the first wall coating for the fusion reactor [4]. In addition, TiC thin films are considered an excellent diffusion barrier between metal silicides and aluminum and are

used in very large scale integration (VLSI) semiconductor technology [5,6].

Most of the TiC thin films have been grown by chemical vapor deposition (CVD) or by some kind of reactive physical vapor deposition (PVD) [2,7]. However, it has been found that relatively high substrate temperature (500 to 1000 °C) is required in order to produce dense TiC films using these methods [2,6,8]. This high substrate temperature can cause serious problems for many applications. For example, it will drastically change the properties of high speed steel substrate due to the change of its microstructure by such high temperature exposure. Film stress caused by different thermal expansion coefficients of the film and substrate may also occur. In addition, unwanted film-substrate interdiffusion and reaction will become unavoidable. Clearly, the development of a low temperature growth process would be helpful to solve these problems.

Pulsed laser deposition (PLD) is an attractive alternative for the deposition of dense TiC at low substrate temperatures [9,10]. Laser induced material removal has been found to occur by at least three mechanisms, depending mostly on photon energy. The first and lowest energy mechanism is laser desorption, where the photons provide only enough energy to desorb weakly bound species from the solid surface. The second mechanism is laser evaporation, where the photons have enough energy to remove chemically bonded species from the surface. In this mode, the laser acts as a thermal source and the irradiated area is in approximate thermodynamic equilibrium. The third and most energetic mechanism is laser ablation. In this mechanism, photons have enough energy to break chemical bonds. Molecular dynamic calculations have described evaporation as local melting of the target, where material is ejected in a broad angular distribution. Ablation, on the other hand, has been described as a more energetic process where material is removed layer by layer, forming well defined pits in the target. Material is ejected in a narrow angular distribution, similar to a supersonic expansion.

Donley et al. have studied the differences in PLD applying different photon energies [11]. Time of flight analyses were conducted using the frequency doubled Nd:YAG laser, λ = 532 nm, and the more energetic 193 and 248 nm radiation available from an excimer laser operating with

ArF and KrF gas, respectively. It was found that, using photons of 532 nm wavelength to remove materials, the velocity distribution of evaporated particles is well described by a Maxwell-Boltzmann distribution with an effective translational temperature of 1500 K, indicating an evaporation mechanism. Using the 248 and 193 nm photons of the excimer for material removal, the velocity distribution of the particles is well described by a drifted Maxwell-Boltzmann distribution, or a supersonic expansion. The effective translational temperature for the 248 and 193 nm wavelength was 63,000 and 111,000 K, respectively, which suggests an ablative mechanism. The average perpendicular velocity of the ablated particles was about 6.5 and 8.0 km/s, respectively. The work proposed here will focus on investigating PLD coatings by the more energetic excimer laser.

The high kinetic energy possessed by the ablated species during PLD may be one of the primary reasons for producing high quality TiC at low temperature. In addition, superior structural properties of the PLD grown films may also result from the presence of ions and other excited species in the ablated particle stream. The additional electronic energy present in such excited species could be a source of non-thermal energy for enhancing atomic migration during film growth.

Deposition of tribological coatings by PLD offers several other advantages. Films can be deposited in less than 10^{-5} Pa of background gases, thus producing high purity films and permitting precise control of dopant concentration. Excellent film adhesion may be achieved due to the energetics of the PLD process. Because photo ablation appears to be a congruent process, complex targets may be used to create complex films. In addition, film properties may be controlled by proper selection of laser parameters, dopant gas, substrate temperature and post laser annealing.

Initial studies

Tribological behaviors of TiC coatings grown by PLD are being presently studied at the Air Force's Wright Laboratory in Wright-Patterson Air Force Base, Ohio. Initial study by

Sessler *et al.* indicated that the PLD films grown at room temperature and 300 °C were stoichiometric and crystalline with grain sizes of 2-10 nm [12]. The films displayed very good adhesion in the wear tests, and their friction coefficients ranged from 0.2 to 0.4 against 440C stainless steel and sapphire balls, which compare well with data of films grown with conventional methods. The wear life of the films deposited on annealed stainless steel was lower than that for films deposited on the tempered stainless steel. Annealed steel gives poor mechanical support to the TiC thin film during tribotests and, subsequently, the plastic deformation of the substrate leads to a shortened wear life of the film. Film deposited at room temperature was inherently harder and showed better wear resistance than the films deposited at 300°C. The study clearly shows that PLD growth of TiC films at room temperature has very high potential application as hard, low wear coatings with excellent adherence.

Results of current study

Supported by AFOSR and Wright Laboratory, I spent nine weeks (from June 5 to August 4) in Dr. Jeff Zabinski's group at Wright Laboratory, WL/MLBT, and continued the effort to explore and understand the tribological properties of PLD grown TiC. Several interesting behaviors have been observed as a result of this research project.

The PLD coatings were deposited on M50 stainless steel substrates with a Lamda Physik LPX 110i excimer laser operating with KrF gas ($\lambda = 248$ nm). The laser beam was directed onto a hot pressed TiC target. The substrates were first rinsed with acetone and methanol, and then they were cleaned in the deposition chamber using a wide beam YAG laser prior to deposition. The substrates were set to rotate during the laser cleaning and deposition processes. TiC coatings were also grown using rf magnetron sputtering. The 440C stainless steel substrates were rinsed with acetone and isopropanol before they were sputter etched in a diffusion pumped MRC 902 in-line sputter deposition chamber. After the sputter etch, the chamber was backfilled with argon and methane at constant flow rates that were kept during deposition. The chambers were pumped to a base pressure of $< 1 \times 10^{-6}$ torr prior to the laser cleaning or sputter etch of the substrates in both PLD and magnetron sputtering cases. The substrates were not heated during deposition.

The nanohardness was studied by a Nano Indenter[R] II of Nano Instruments, Inc., which allows simultaneous determination of thin coating hardness and elastic modulus. A typical procedure in a hardness/modulus experiment consisted of the following steps: (1) Approach; (2) Loading; (3) Partial unloading; (4) Hold; (5) Loading again; (6) Hold; and (7) Final unloading. The hardness and elastic modulus were calculated from the final unloading curve after the effects of thermal drift and creep were taken into account.
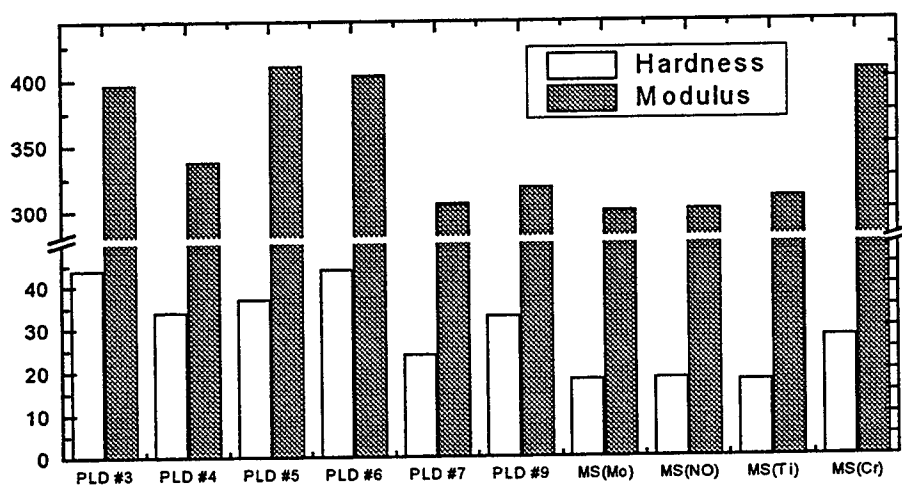

Figure 1. hardness and elastic modulus of six PLD coatings and four magnetron sputtered (MS) coatings

The hardness and elastic modulus of six PLD coatings and four magnetron sputtered (denoted as MS) coatings are shown in Fig. 1. PLD#3 and PLD#6 were two nearly identical coatings of 8000Å thick. PLD#4 seemed to contain excess oxygen and had a gold color appearance. PLD#5 is a sample similar to PLD#3 and PLD#6, except that it was about 2 μm thick PLD#7 was a sample PLD deposited with *in situ* YAG laser annealing. A YAG laser was directed onto the coating to anneal it while it was grown by the excimer laser. PLD#9 was PLD deposited while applying excimer laser *in situ* annealing. During the growth, the excimer laser beam was split into two, one directed at the target and the other directed at the coating. As seen

in Fig. 1, the hardness of the last two PLD coatings (PLD#7 and PLD#9) was reduced as a result of the *in situ* laser annealing. The hardness of the PLD TiC coatings deposited at room temperature without laser annealing was as high as 44 GPa (samples PLD#3 and PLD#6). In contrast, the hardness of magnetron sputtered TiC coatings deposited at room temperature had values near 20 GPa, *i.e.*, only half of that of PLD films. Thus it is obvious that the PLD TiC coatings grown at room temperature were much harder than the TiC films grown by magnetron sputtering under the given experimental conditions. This demonstrates again the ability of PLD to grow hard and dense TiC at low temperature. The elastic modulus, which reflects more intrinsic properties of crystal lattice, correlated well with the hardness.

Scratch test, which determines the adhesion of coating to the substrate, was conducted on a CSEM scratch tester. In the test, the critical applied load to cause the coating to detach from the substrate was determined, and the detachment of the coating was monitored through acoustic emission signal. The experiments conducted on the PLD films and magnetron sputtered films indicated that the softer magnetron sputtered TiC coatings adhere better to the stainless steel substrate than the PLD grown coatings (see Fig. 2 and compare samples MS with PLD). In Fig. 2, the adhesion is measured as the critical load to cause the coatings to delaminate. The relatively poor adhesion of the PLD coatings obtained from this scratch test is probably the result of its high hardness and brittleness, since softer and tougher materials absorb the identical applied load better through plastic deformation.

Figure 3 shows the scares on sample PLD#3 at the early stage of a scratch test when the normal load was small and no coating penetration had occurred. The cracks shown are typical of those coatings of high hardness and brittleness.

In a parallel study, we have examined the adhesion of magnetron sputtered TiC coatings when a metallic interlayer is inserted between the coating and stainless steel substrate. Scratch tests were conducted on four different magnetron sputtered TiC coatings. One had a Ti interlayer, denoted as MS(Ti); one had a Cr interlayer, MS(Cr); and one had a Mo interlayer, MS(Mo). The
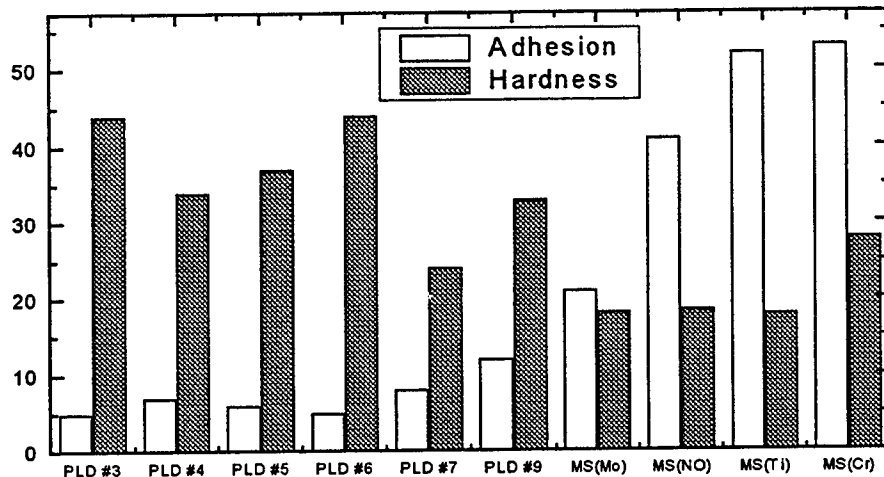
Figure 2. adhesion and hardness of six PLD coatings and four magnetron sputtered (MS) coatings.

fourth one contained no interlayers and TiC was directly deposited on the stainless steel substrate, MS(NO). All interlayers had the same thickness (1000Å) and the TiC coatings were 3 μm thick. Other deposition parameters were chosen the same. As seen in Fig. 2, Mo interlayer has a detrimental effect on the adhesion caused probably by the poor stress bearing capability of the porous Mo film deposited at low temperature [13]. However, it is evident from the figure that the insertion of both Ti and Cr interlayers enhances the adhesion of TiC, by as much as 25%. One may note that in the case of Cr interlayer insertion the hardness and modulus of the TiC coating had also higher values than those when other interlayers were used. We are currently studying the causes for these enhancements and the effects of interlayer thickness on the adhesion of TiC coatings to the stainless steel substrates.

The coefficient of friction of the PLD films measured with an Implant Sciences Corp. pin-on-disk type tribometer had a typical value of about 0.2 - 0.3 when using a 440C stainless steel pin, which is consistent with the data obtained previously by the Wright Laboratory group. The coefficients of friction of the magnetron sputtered TiC coatings were comparable to that of the PLD ones. The friction curve of magnetron sputtered TiC without an interlayer, MS(NO), is
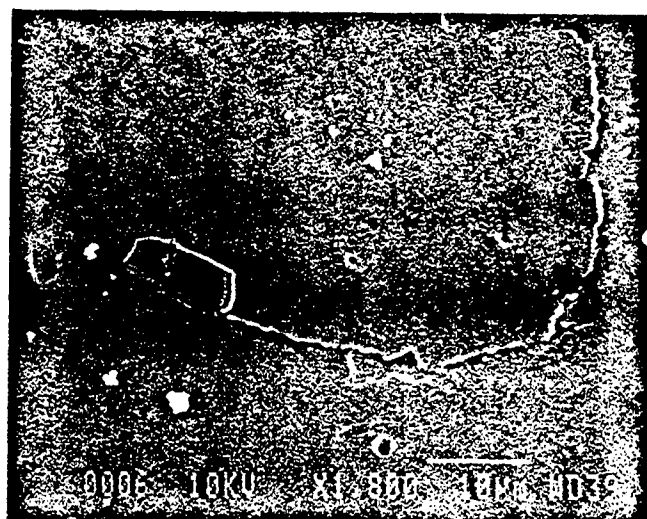
shown in Fig. 4 as a function of sliding cycle.

Figure 3. Scares on sample PLD#3 at the early stage of a scratch test when the normal load was small and no coating penetration had occurred.
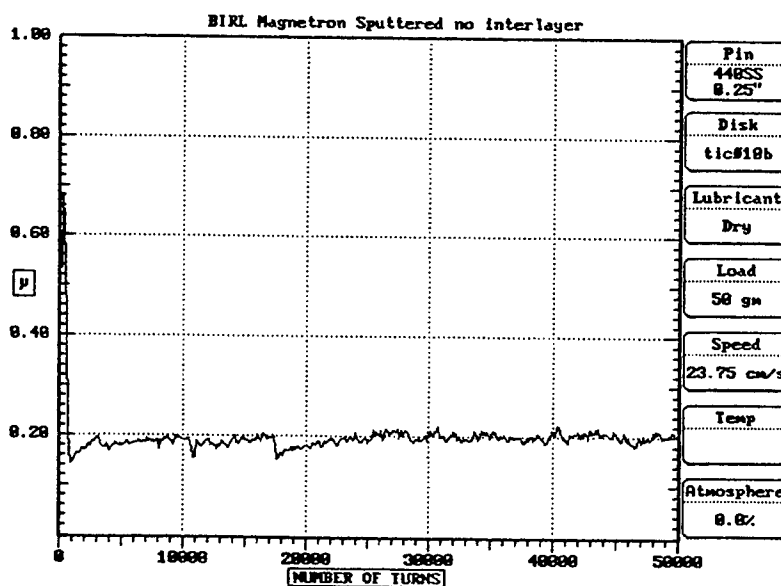

Figure 4. Friction curve of magnetron sputtered TiC without an interlayer, MS(NO), as a function of sliding cycle.
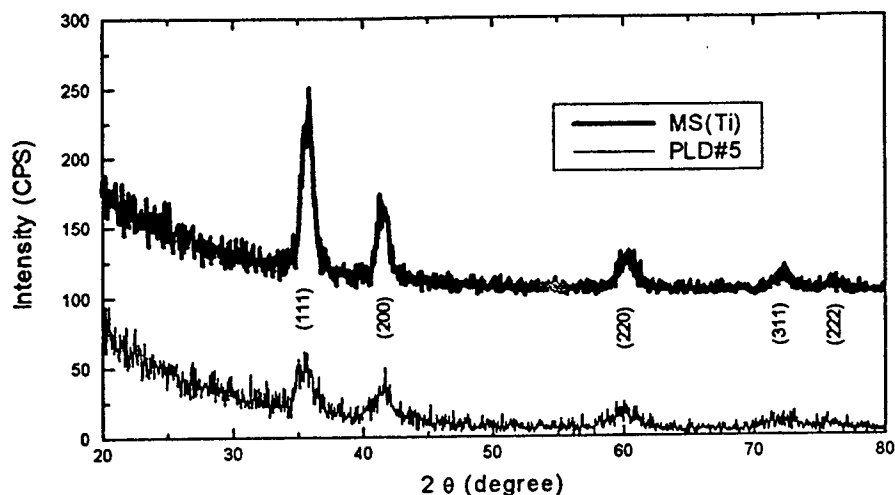
54-10

Figure 5. X-ray diffraction patterns of samples MS(Ti) and PLD#5.

Figure 5 shows the x-ray diffraction patterns of samples MS(Ti) and PLD#5. The incident angle θ was fixed at 0.4 degree and 2θ was scanned from 20 to 80 degrees. As can be seen, both magnetron sputtered and PLD coatings consist of primarily fcc TiC. From the peak width, the grain sizes of both MS and PLD coatings were estimated to be close to 8 nm.

The cross section of the PLD#3 was examined by SEM. A SEM micrograph is shown in Figure 6. The layered structure of the coating is evident from the micrograph. The substrate was M50 stainless steel.
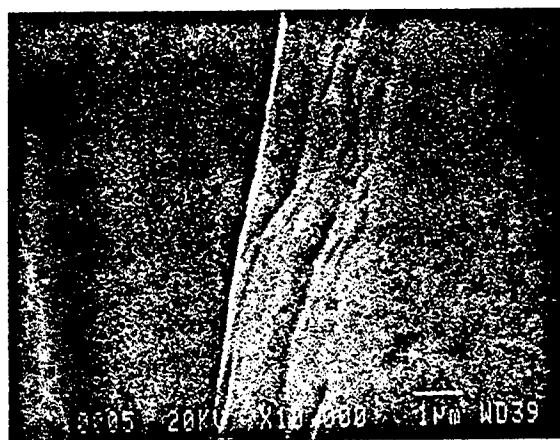

Figure 6. SEM micrograph of the cross section of PLD#3 coating.

Future Studies

Based on the results discussed above, we plan to pursue four major approaches to improve the tribological properties of the PLD grown TiC coatings: (a) applying Cr or Ti interlayer to the PLD TiC coatings to enhance adhesion; (b) using post ion implantation to modify the interface; © in-situ laser annealing of the coatings; and (d) investigating TiC/TiN multilayered coatings.

(a) Cr or Ti Interlayer

As mentioned in the above discussion, the adhesion of PLD TiC coatings compare poorly with the magnetron sputtered ones. This might be attributed to the fact that PLD grown TiC is very hard and brittle. Another contributing factor for the adhesion problem is obviously that the coatings were deposited at low temperature (this mechanism applies also to our magnetron sputtered coatings). Low deposition temperature may cause problems arising from condensation of undesirable vapors, low reaction rate of chemical process and low diffusivity of the elements.

The first approach to address the adhesion problem is to include a Cr or Ti interlayer between PLD TiC coating and substrate. As demonstrated in our studies, Cr or Ti interlayer seems to increase the adhesion between magnetron sputtered TiC coatings and stainless steel substrates. One possible reason is that Cr or Ti are both capable of forming hard carbides with strong bonds and therefore may improve the adhesion. The adhesion may also be promoted by the presence of metal oxides at the interface [14,15].

(b) Ion Implantation

It has been shown that enhanced adhesion between metallic thin films and ceramic substrates can be obtained using high energy ion implantation. It is believed that MeV ion-induced diffusion as well as interface chemistry can both be important factors in adhesion enhancement [16]. Both effects occur simultaneously under MeV ion bombardment so that chemical effects will probably become more pronounced at low temperature.

It is proposed that the PLD grown TiC coatings, with or without metal interlayers, be

subjected to post high energy ion implantation. This part of the project will be conducted in collaboration with Dr. Gary Glass at the University of Southwestern Louisiana, whose laboratory is equipped with a 1.7 MV National Electrostatics 5SDH-2 Tandem Pelletron^R accelerator system capable of producing various ions with energies 0.3 - 8.0 MeV.

(c) *In Situ* Laser Annealing of the Coatings

By annealing the coating as well as substrate *in situ*, the structures of the coatings and substrate will be modified to promote better adhesion. The chemical reaction and diffusion at the interface may also be positively affected by the laser annealing. Both Nd:YAG and excimer lasers will be used as the annealing thermal source in order to identify the ideal photon energy for such annealing.

(d) TiC/TiN multilayer coatings

TiN, although not as hard as TiC, has much better toughness and is less brittle that TiC. It is natural to combine the two in coating applications. In fact, TiC/TiN multilayer coating has been studies extensively [17,18], and results show better overall performance of the coatings, including stress handling. However, no such study has ever been done using PLD process and it is planned that TiC/TiN multilayered coatings be investigated in this project. In making the coating, the target used will be TiC, and N_2 gas will be periodically introduced into the chamber. The thickness of each TiC or TiN layer will be controlled and optimized.

It is hoped that the four approaches mentioned above will not only improve the coating adhesion, but other tribological properties (wear, friction, hardness and toughness) as well.

References

[1]   M. Sjostrand, *Met. Powder Rep.*, **41**, 905 (1986).

[2]   G. Georgiev, N. Feschiev, D. Popov and Z. Uzuuov, *Vaccum*, **36**, 595 (1986).

[3]   P. K. Ashwini, V. Kumar and S. K. Sarkar, *J. Vacuum Sci. Technol. A*, **7**, 1488 (1989).

[4]   A. E. Kaloyeras, W. S. Williams, F. C. Brown, A. E. Greene and J. B. Woodhouse,

Phys. Rev. B, **37**, 771 (1988).

[5]   A. Appelbaum and S. P. Murarka, *J. Vacuum Sci. Technol. A,* **4**, 637 (1986).

[6]   M. Eizenberg and S. P. Murarka, *J. Appl. Phys.,* **54**, 3190 (1983).

[7]   J. F. Sundgren, B. O. Johansson and S. E. Karlsson, *Thin Solid Films,* **105**, 353 (1983).

[8]   J. E. Sundgren, B. O. Johansson, H. T. G. Hentzell and S. E. Karlsson, *Thin Solid Films,* **105**, 385 (1983).

[9]   O. Rist and P. T. Murray, *Mater. Lett.,* **10**, 323 (1991).

[10]   M. S. Donley, J. S. Zabinski, W. J. Sessler, V. J. Dyhouse, S. D. Walck and N. T. McDevitt, *Mater. Res. Symp. Proc.,* **236**, 461 (1992).

[11]   M. S. Donley, J. S. Zabinski, V. J. Dyhouse, P. J. John, P. T. Murray and N. T. McDevitt, *Lecture Notes on Phys.,* **389**, 271 (1991).

[12]   W. J. Sessler, M. S. Donley, J. S. Zabinski, S. D. Walck and V. J. Dyhouse, *Surface and Coatings Technol.,* **56**, 125 (1993).

[13]   R. A. Hoffman, J. C. Lin, J. P. Chambers, *Thin Solid Films,* **206**, 230 (1991).

[14]   M. Van Stappen, B. Mallliet, L. De Schepper, L. M. Stals, J. P. Celis and J. R. Roos, *Surf. Eng.,* **4**, 305 (1989).

[15]   M. D. Bentzon, K. Mogensen, J. Bindslev Hansen, C. Barholm-Hansen, C. Træholt, P. Holiday and S. S. Eskildsen, *Surf. and Coatings Technol.,* **68/69**, 651 (1994).

[16]   J. P. Celis, M. Franck, J. R. Roos, E. W. Kreutz, A. Gasser, M. Wehner, K. Wissenhach and N. Pattyn, *Appl. Surf. Sci.,* **54**, 322 (1992).

[17]   E. Vancoille, J. P. Celis and J. R. Roos, *Tribology International,* **26**, 115 (1993).

[18]   S. Eroglu and B. Gallois, *Surf. And Coatings Technol.,* **49**, 275 (1991).

Acknowledgments:

I wish to thank the support of AFOSR and Wright Laboratory, especially that of Dr. Jeff Zabinski.

PHOTOLUMINESCENCE STUDIES OF THE RIGID ROD POLYMER
Poly (p-phenylenebenzobisthiazole)(PBZT)

Barney E. Taylor
Visiting Assistant Professor
Department of Physics

Miami University - Hamilton
1601 Peck Blvd.
Hamilton, OH 45011

PHOTOLUMINESCENCE STUDIES OF THE RIGID ROD POLYMER
Poly (p-phenylenebenzobisthiazole)(PBZT)

Barney E. Taylor
Visiting Assistant Professor
Department of Physics
Miami University - Hamilton

Poly (p-phenylenebenzobisthiazole)(PBZT) is a rigid rod polymer with excellent thermal and mechanical properties. A previous study characterized electroluminescent devices with PBZT as the active layer fabricated upon an indium-tin oxide (ITO) coated glass slides. In the previous study, the spectral studies showed an emission from 850 nm (the long wavelength limit of the apparatus) up through about 525 (nm) at device voltages ranging from 2.35 V to 4.0 V for various devices studied. This report describes a systematic investigation of the photoluminescence (PL) of PBZT to complement the electroluminescent studies performed earlier.

A number of different PBZT samples were studied by room temperature photoluminescence. A systematic study of the PL of one sample was performed over the range of -163 °C to 291 °C. Analysis of the temperature dependence of the PL spectra indicate a thermally activated process with an activation energy of 83 meV. Some structure was observed in the PL spectrum of PBZT, and a process of fitting multiple gaussian lineshapes to identify the relative intensity and width of each component was undertaken. Very consistent fitting results were obtained. The physical significance of the variation of the components with temperature is not immediately evident and further studies are required.

The relatively low electroluminescent yield of the PBZT devices in the earlier study has led to a desire fabricate electroluminescent devices based upon other polymers that, hopefully, have a higher quantum efficiency. PL appears to be very useful as a screen for likely polymer EL candidates.

PHOTOLUMINESCENCE STUDIES OF THE RIGID ROD POLYMER
Poly (p-phenylenebenzobisthiazole)(PBZT)

Barney E. Taylor

Introduction

There is much interest in electroluminescent polymeric devices for use in displays and other indicator roles.[1] At this point in the evolution of electroluminescent devices, quite impressive brightness has been achieved in prototype devices. However, short device lifetimes of the device are also the rule rather than the exception.[2] Presumably, rigid rod polymers offer hope of longer device lifetime due to their superior mechanical and thermal properties.[3] An earlier study focused on PBZT as the active layer.[4] Electroluminescence was observed, but with a very low quantum efficiency. A spectrum was observed that extended from about 520 nm to the experimental limit of 850 nm. The thrust of this investigation was to obtain the complementary photoluminescence (PL) spectrum from the same samples and to compare the EL and PL spectra. In most materials, the PL and EL spectra are essentially similar. This was not the case for PBZT. At this time it is not clear whether the EL is due to some indirect, phonon mediated, mechanism or the result of mid level traps. The apparent difference in the spectra that were expected to be essentially similar has lead to a more thorough investigation of the PL of PBZT than had been initially planned. The structure of PBZT is shown in Figure 1.

During the interval since the previous EL investigation, an optical multichannel analyzer (OMA) has been obtained by the Physics Group within the Polymer Branch of Wright Laboratories Materials Laboratory. This new instrumentation allows significantly extended experimental capability over the previous experimental system. Attempts to extend the EL measurements of the earlier study were largely frustrating do to low device yield.[5] Some spectra were taken, but they did not readily correspond to the spectra of the earlier EL study. Hence, spectral normalization of the new apparatus was a prime interest, so that true spectral profiles could be obtained. A significant portion of this effort has been devoted to refining the spectral normalization of the data obtained from the OMA.

The low yield of fabricating successful PBZT EL devices has been driving the search for other polymeric systems that might be suitable candidates for successful, long lived, EL devices. Due to the significant overhead involved with fabricating prototype EL devices, it was decided to see if PL could be used as an initial screen. Devices with low PL yields are expected to also exhibit low EL yields. One material 6F-PBO - COPY, a co-polymer blend, was surveyed by PL to check the viability of such screening. Room temperature PL results showed a more intense luminescence than that of PBZT. The PL occurred at a higher energy than PBZT, offering hope for devices beyond the red region of the visible spectra. At this time, the PL results are preliminary, and no devices have been fabricated upon this material. But this system and several related polymers are being actively considered for fabrication of prototype EL devices.

One result of the PL study was the discovery of structure in the PL spectra -- even after spectral normalization. Significant time has been spent fitting multiple gaussian lineshapes to the PL spectra of PBZT. In general, good fitting results were obtained from the fits. The best fit parameters yield the intensity of the line, its width and central energy. Since no specific theoretical explanation of the photoluminescence of PBZT exists in the literature, the usual kinds of empirical relationships -- Arrhenius, power law, exponential, etc. -- have been investigated. With the exception of the integrated area, no strong trends were determined.

EXPERIMENTAL

Sample Fabrication and Mounting

Most of the experiments were performed on PBZT films cast upon indium-tin oxide (ITO) coated glass by a spin coating technique.[6] The spin coated films were typically 2.5 cm square and 0.1 micron thick. Some of the samples had been previously used for EL studies and had metallic electrodes evaporated onto the PBZT. Careful alignment minimized the amount of electrode covered area sampled. The presence of the electrodes did not seem to alter the shape of the PL spectrum obtained, although the intensity was affected. Other samples were in the as-received condition. Samples were stored in plastic containers under ambient atmosphere, since PBZT is relatively immune to atmospheric moisture and oxygen.

Room temperature samples were mounted in an optical filter mount and studied. Temperature dependent studies were performed with the sample mounted in a Specac™ cold finger cryostat. The sample was mounted on the cold head, and held in place with a spring steel screen at the bottom to ensure good thermal contact. The cryostat was evacuated to a few millitorr using a rotary forepump. Vacuum levels were monitored with a thermocouple vacuum gauge. The Specac cryostat has a useful temperature range from about -165 $^{\circ}$C to about 300 $^{\circ}$C. The Specac is not truly vacuum tight, it leaks up to a few Torr over the course of a temperature dependence run, but the heat leakage is sufficiently minimized to allow successful completion of the experiments.

Optical Instrumentation:

A PAR 4000 series OMA was used for the studies being described. The 4000 system consists of a quarter-meter monochromator, with fixed input slits (25 microns), attached to a cryogenically cooled (-120 $^{\circ}$C) 1024 by 256 element CCD. While not intensified, the CCD is capable of long integration times with extremely low noise when at its operating temperature. A dedicated 486-66 class computer housed the interface card for the OMA and served as the control center of all measurements made on the OMA. The 4000 series OMA has extremely capable software for acquiring and storing spectra over a wide range of experimental conditions.

The monochromator on the OMA was equipped with three gratings mounted on a turret. Each grating has 150 grooves per millimeter. The three gratings are blazed at 300 nm, 500 nm and 800 nm for optimum response in the near UV, the green and the red regions of the spectrum.

Excitation for the PL experiments was obtained by an Acton™ quarter-meter monochromator equipped with a 3 grating turret, holding gratings identical to those in the OMA. Light energy is provided by either a deuterium lamp or a 45 watt tungsten halogen lamp. The Acton spectrometer is fitted with adjustable entrance and exit slits to allow exchange of narrow spectral bandwith for more excitation beam intensity. In most cases slit widths of 1 mm were used on the Acton monochromator resulting in a excitation beam dispersion of 12.8 nm. The PL spectra were quite wide in comparison.

Wavelength calibration of both instruments was performed using He-Ne lasers and lines from a mercury pencil lamp. An Oriel Calibrated lamp was used for spectral normalization purposes.

Experimental Configuration:

Figure 2 shows a block diagram of the photoluminescence experimental system. Light, usually from the tungsten source, passed through the Acton monochromator and fell upon a concave mirror. The mirror focused the slit image into a point on the sample. The inclination of the sample with respect to the beam is such that the specular reflection was directed away from the second mirror that picked up the emitted luminescence. During the alignment process tradeoffs were made between the optimum orientation of the sample for collecting the emitted luminescence and for directing the specular reflection away from the OMA to prevent ghosting or other problems. The light collected by the second mirror was focused onto the fixed entrance slit of the OMA and ultimately fell upon the CCD for detection.

Optical filters were used in two locations: Bandpass filters at the exit slit of the monochromator cleaned up any optical deficiencies in the exit beam due to scattered light, etc. A long pass filter at the entrance slit to the monochromator removed most of the low level component of the probing wavelength to prevent second order effects of the excitation radiation from appearing in the PL spectra. The filters had already been optimized for excitation beams in the 400 to 430 nm range and luminescence in the 450 to 900 nm range.[5] A GG5 filter was used at the entrance of the OMA, while BG3 and BG40 filters were used at the exit slit of the Acton monochromator. In order to insure that the light was entering directly on the axis of the OMA, the collection mirror was mounted on a track that was fixed along the optical axis of the PAR monochromator.

Stray light was minimized by a baffle constructed of black cloth. The excitation wavelength, grating and lamp source were selected manually by a keypad for the Acton monochromator. The settings of the monochromator on the OMA were controlled through the GPIB interface by the dedicated computer.

Integration times of 64 seconds were typical for the PL measurements, with weak, or poorly aligned samples requiring even more time. Calibration standards required much, much shorter integration times -- typically 100 to

400 ms. The OMA records all of the settings of the OMA with each set of data saved, allowing one complete knowledge of that part of the experimental conditions. Other pertinent information such as the width of the slits on the Acton, the excitation wavelength and the filters being used were recorded as a comment in the OMA data file.

When temperature dependence runs were being performed, there was a nearly continual need to replenish the liquid nitrogen in the cold finger. This presented some obstacles -- it was extremely easy to slightly bump the cryostat and alter the optical alignment during the refilling process. Special care was taken to avoid disturbing the Specac cryostat. If bumping was suspected, a second spectrum was taken for comparison before the setpoint temperature was altered.

Data and Analysis:

Figure 3 is a room temperature, PL spectrum of an as-received PBZT excited with a excitation wavelength of 400 nm. The circles represent the raw data, the dashed line the spectrally normalized data and the inset graph is the multiplicative spectral normalization function. The spectrum is typical of the as-received samples. The spin coated PBZT films were reasonably uniform, although different colorations, indicating different thickness, were visible near the edge of the sample. The excitation beam was focused into a slit image on the sample and probed both the central uniform region and the smaller nonuniform regions near the top and bottom of the slit image. When the excitation beam was moved to a different location on the sample, a slightly different spectra was often observed -- both in intensity and, to a lesser extent, in spectral shape. Attempts to acquire PL spectra from the bare ITO coated glass showed no luminescence.

There are at least two features present in the spectrum of Figure 3, one is around 550 nm and other about 650 nm. The shape of the spectrum of Figure 3 is not in good agreement with the PL spectra of PBZT found in the literature. [7]

The PBZT was dissolved in MSA, an acid with a low vapor pressure, in order to be spin coated. It was postulated that perhaps the different shape of the spectrum in Figure 3 compared to the literature was related to residual MSA. The solid line in Figure 4 shows the raw, room temperature, spectrum of

a sample from the same batch that has been washed in ethanol for several weeks. This spectrum shows an enhancement of the higher energy (550 nm) peak relative to Figure 3. The shape of the spectrum in Figure 4 is much closer to the shape reported in the literature for PBZT. [7, 8]

There is a background level of about 1000-1100 counts in the raw data of Figure 4. Most of the background is intrinsic to the PAR OMA. The remainder is due to stray light in the laboratory from LED's, the CRT screen and other sources. A background spectrum was usually taken for each sample. Hence, the true, unnormalized PL spectrum must have the background subtracted from the raw data. For spectra with a good signal to noise ratio, this amounts to just shifting the baseline. For very marginal spectra, the gentle curvature of the background can affect the shape of the raw and spectrally.

An Oriel Calibrated lamp was used for spectral normalization purposes. The lamp was placed relatively distant from the OMA, and allowed to stabilize. The same second order blocking filter was placed at the entrance slit of the OMA as was used when acquiring PL data. A set of spectra were taken (at much shorter integration times -- to prevent overloading the CCD), along with background spectra with the lamp off. A multiplicative function was calculated by the following equation and a computer program was generated to perform the normalization.

$$PL_{norm} = (PL_{raw} - BG) \frac{Lamp\ Function}{OMA_{lamp} - BG_{lamp}}$$

where BG is the background associated with a given set of data, OMA_{lamp} is the data obtained for the calibrated lamp and BG_{lamp} is the background for the calibrated lamp. Lamp Function is the energy data supplied by Oriel for the spectral luminosity of the lamp. Hence, the normalization is relative to incident energy rather than the nubmer of photons. The computer program was then checked against other spectra from the normalizing set to ensure a proper shape. The experimental spectra were background subtracted and the passed through the normalizing program, resulting in spectrally normalized data files that were used for nonlinear least squares fitting. The resulting curve is the broken line in Figure 4.

Since the ethanol washed PBZT sample yielded a spectrum that was in better agreement with the literature than the unwashed sample, it was used for a series of PL measurements. The goal was to try to fully characterize the PL

response of the ethanol washed PBZT prior to fabricating EL devices on that sample. One of the tests was an excitation spectrum. The OMA system is capable of obtaining excitation data, but that requires a significant number of scans, and much data massaging to construct an excitation spectrum from the spectral data. An excitation spectrum was graciously performed by Dr. Natarajan of MLPJ using a Perkin Elmer Spectrofluorimeter and the resulting spectrum is shown in Figure 5, along with the absorption of a PBZT sample cast from solution in order to be thick enough for absorption measurements. In general the excitation spectrum should mirror the absorption of a given PL sample. There are similarities in the two spectra, but they are far from identical. It is not known if the differences are related to the difference in thickness or the method of casting. The tail at about 600 nm in the excitation is an artifact of the instrument used to acquire the excitation.

A temperature dependence study on the PL of the ethanol washed sample was performed. A complete set of PL spectra was obtained in one session, from temperatures of -163 °C to 291 °C. One thing of interest was how the intensity and shape of the PL spectrum would be affected by temperature changes. The spectra were background subtracted, spectrally normalized and then scaled to unity maximum amplitude. A few of the resulting family of curves are shown in Figure 6. Figure 6 shows a broadening of the spectrum, and a shift to lower energy as the temperature of the sample is increased. The integrated area of the PL spectra have been plotted against temperature in Figure 7. The total PL was approximately constant for low temperatures, below 175 K, and fell off as the temperature increased. Pankove [9] describes the process as a reduction in the quantum efficiency as the temperature increases, due to thermally activated nonradiative recombination sites. The integrated area and the temperature should obey the following relationship:

$$\eta(T) = \frac{1}{1 + C\,e^{-E^*/kT}} = \frac{PL(T)}{PL(0)} \quad ,$$

where η is the quantum efficiency and E^* is the thermal activation energy. Since the T = 0 K PL must be known to calculate η, a slightly modified form was adopted for nonlinear least squares fitting:

$$PL_T = \frac{A}{1 + C\,e^{-\frac{T_a}{T}}} \quad ,$$

where A is effectively the integrated area at T = 0, and T_0 is the equivalent temperature of the thermal energy. Hence, the activation energy is given by $E^* = k_b T_0$. Figure 8 is a plot of the integrated PL intensity versus inverse temperature. The experimental data points are shown by the symbols and the best fit results are plotted as the solid line. The scatter in the low temperature data is most likely due to slight alterations of the alignment during the continual refilling of the liquid nitrogen in the cold finger. The activation energy obtained from the nonlinear least squares fit is 83 meV. An alternative explanation of the 83 meV activation energy is the ionization of defect levels near the band edge as described by Brown [10].

The structure in the peaks was investigated by deconvoluting the spectrally normalized data into multiple gaussian peaks. Exceptionally good results could be obtained as indicated by Figure 9, which is a plot of the raw data (marking only some of the data points), the composite fit and the individual components. The results of the fitting of the temperature study are shown in Table 1. The trend in the data is for a decrease with temperature of the 2.2 eV peak, with the other three peaks exhibiting modest increases in energy up to about 500 K. The width of the 2.0 eV peak increased with temperature, showing a dramatic rise above 450 K. The width of the 2.05 eV peak exhibited a similar behavior. While, the width of the 2.4 eV peak decreased slightly, the width of the 2.2 eV peak was almost constant. All of the amplitudes exhibit a drop with increasing temperature up to about 450 K. Above that temperature, the 2.2 and 2.4 eV peaks show a growing amplitude.

Discussion:

There is no doubt that washing the spin cast PBZT film in ethanol had a profound effect on the photoluminescence, yielding a shape that is qualitatively similar to that reported by Jenekhe. [7, 8] Unfortunately no other unused samples of the spin coated PBZT exist to verify the repro-ducibility of the effect caused by the ethanol wash. There are multiple features in both the as-received and the ethanol washed PL spectra, and the washing seemed to shift the relative concentration of the peaks. The effect of washing is long lived. PL measurements on the sample several months after being removed from the wash yielded essentially the same spectra. One

possible explanation for the change is that residual MSA in the as-received sample was deprotonating the polymer, and the long wash in ethanol lead to the diffusion of the MSA out of the sample. However, Lefkowitz [11] has studied the effects of deprotonation in the PBO system, and found that deprotonation has little effect on the PL obtained compared to that associated with aggregation. Other PL studies of PBZT within the physics group at MLBP [5] have demonstrated interference artifacts in the PL spectra of doctor-bladed and freestanding PBZT films. The interference is due to thin film effects as described in any introductory physics text. It is possible that some of the variation of the shape of the PL spectrum with position of the excitation beam on the sample was due to nonuniformity of the thickness of the sample.

The temperature dependence data is quite interesting. We see that the total PL intensity decreases with increasing temperature. The explanation in terms of the quantum efficiency seems quite valid considering the success of the fit to the experimental data. The activation energy for the nonradiative recombination is 83 meV, which is a plausible energy level. At this time, we can offer no explanation as to the physical origin within the PBZT chain for such thermally activated nonradiative sites. The sites must be attributed to unnamed 'traps' of the free carriers.

Although the quality of the fitting of multiple gaussian peaks to the spectra gives extremely good overall fits, one must be cautious in placing too much faith in the parameters generated. In general nonlinear least squares fitting excels when working with sharp functions, with well defined extrema and inflection points. With the exception of the 2.4 eV peak, the shape of the PL spectra of PBZT is relatively smooth. This leads to a very shallow minima in χ^2 space, which means that the fitting algorithm can wander substantially while settling in upon a minimum. This behavior was observed during the initial fit -- the value of χ^2 would stabilize, and then change very slowly in succeeding iterations. The user could watch the present value of the parameters and see changes in two peaks occurring during an iteration -- one would have a parameter such as its energy decrease, and the other would have a compensating increase. The increase in the width of the 2.2 eV peak is definitely significant. As the unity scaled spectra of Figure 6 show, the overall width of the spectrum is increasing at elevated temperatures. The overall width is greatest in the 2.0 eV peak and the variation with

temperature is the greatest. Thus the predominant factor in the greater width (in proportion to height) of the PL spectra of PBZT with increasing temperature is the increase in the 2.0 eV feature.

One may question the use of gaussian fitting functions, rather than lorentzians for the PL data of PBZT. Indeed, the existence of both a 2.0 and 2.05 eV peak suggest that the shape might be lorentzian instead of gaussian. During the preparation of this report, some of the data was refit with multiple lorentzian lines. Instead of the 4 peaks needed for gaussian fits, 5 lorentzian peaks were needed to fully replicate the experimental lineshape. Figure 10 shows the result of fitting lorentizn lineshapes to the same set of data as Figure 9.

Comparison of the best fit parameters for both types of fitting for the 133 K run shows that the 2.4 eV peak is comparable, while the 2.18 eV peak for the gaussian has been replaced by two lorentzian peaks, at 2.16 and 2.21 eV. The 2.03 and 1.99 eV gaussian peaks have a single wider lorentzian counterpart at 2.02 eV. The remaining lorentzian fit, which has the greatest width of any of the peaks, has its peak centered at 1.86 eV. The additive constant that compensates for residual baseline is smaller in the case of the gaussian than for the lorentzian fit. Above 450 K, the additive constant is significantly smaller for the lorentzian fits than for the gaussian. This is quite unexpected. Thermal broadening of the PL peaks should increase as the temperature increases and more thermal energy is available, causing an, assumed, lorentzian curve to broaden into a regime that is better described by a gaussian.

Conclusion and Suggestions for Further Work:

The effect of washing the spin coated sample in ethanol was very interesting. It is essential that an attempt be made to duplicate the results. If the results can be duplicated, then it would seem reasonable to perform quantitative and qualitative analysis on the PBZT material (or the ethanol) in an effort to determine what was altered during the washing process. Other questions worthy of interest have to do with the choice of ethanol. Would methanol, isopropyl or, even, distilled water have worked as well?

The PL spectrum of PBZT is very interesting due to the multiple peaks present and their variation with sample treatment. There is no doubt that the total integrated area of the PL spectrum is dominated by a thermally activated nonradiative recombination center at higher temperatures. The activation energy was found to be 83 meV, which is a reasonable number. It seems crucial that a second temperature study be performed upon as-received spin cast PBZT. The goal would be to look for similar behavior and to compare the activation energy with that of the washed sample.

By performing PL studies at or near liquid helium temperatures, it is possible that the multiple peaks would further narrow and yield a spectrum with much greater structure. Even if the individual peaks were not fully resolved, the increased structure would greatly increase the slope around the minima in χ^2-space, leading to greater confidence in the best fit parameters compared to those of the present study.

At the time of preparation of this report, the washed PBZT sample was being prepared for EL measurements. If the EL experiments are successful, the results of those measurements could possibly confirm the shape of the EL obtained from the doctor-bladed samples of the previous study [4] and extend the near infrared portion of the EL spectrum beyond the 850 nm limit of the previous study. That would indicate that the different way of forming the film and the ethanol wash did not alter the basic EL nature of PBZT. If however, the EL spectra should be significantly different from that of the previous study, it would raise the question of whether the different method of forming the film -- i.e., doctor-blade versus spin casting, or the ethanol wash was responsible for the difference. New samples would need to be prepared and an exhaustive study performed.

The preliminary measurements on the 6fPBO - COPY sample seem to indicate the merit of using PL as one of the screening tests for viability of potential EL materials. Coupled with conductivity studies to insure that the samples can carry sufficient current to form useful devices, the PL screening can guide the choice of material to be used as EL candidates.

The luminescence of PBZT, both photo and electro, is a very rich area for study. Thoughtful and creative experiments and a thorough theoretical treatment are needed to be able to fully understand this interesting system.

Acknowledgments:

References:

1   J. Kido "Organic Electroluminescent Devices Based on Polymeric Materials", Trends in Polymer Science, **2**, 10 (1994).

2   S. Forrest, P. Burrows, and M. Thompson, "Organic emitters promise a new generation of displays", Laser Focus World, 99, Feb., 1995.

3   J. Wolfe, B. Loo, and F. Arnold, Macromolecules, **14**, 915, (1981).

4   B. E. Taylor, Final Report, AFOSR SFRP Program, Sept., 1994.

5.   J. B. Ferguson, Private Communication.

6   Lora A. Cintavey, Master's Thesis, U. of Cincinnati, 1995.

7   S. Jenekhe and J. Osaheni, Science, **265**, 765, (1994).

8   J. Osaheni, S. Jenekhe, and J. Perlstein, J. Phys. Chem., **98**, 12727 (1994).

9   J. Pankove, Optical Processes in Semiconductors, p. 166, Dover, 1971.

10   F. Brown, Physics of Solids, pp. 366-370, W. A. Benjamin, 1967.

11   S. Lefkowitz and D. Roitman, Polymer, **35**, 1576, (1994).
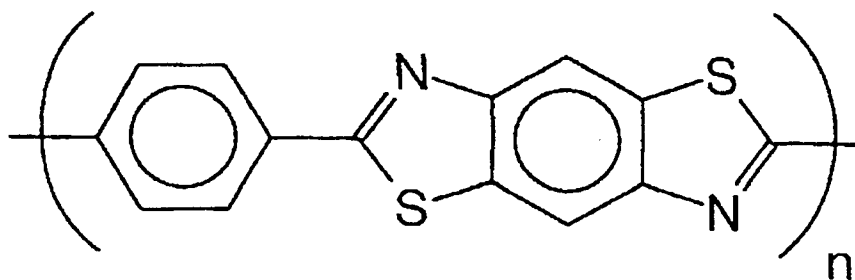
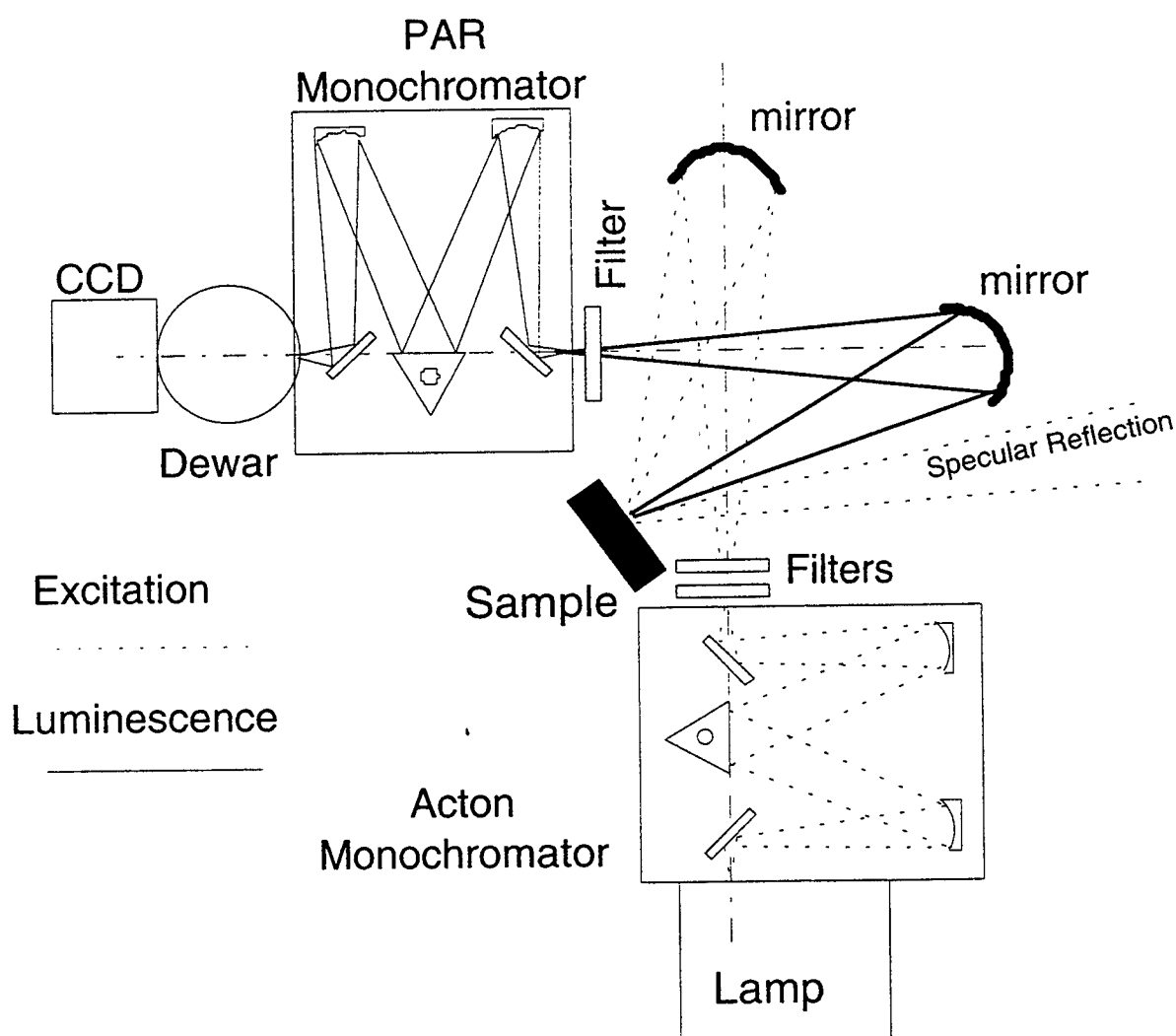Figure 1. Structure of PBZT.

Figure 2. Optical Schematic of the PL apparatus. The excitation is shown in broken lines and the luminescence is shown in solid lines.

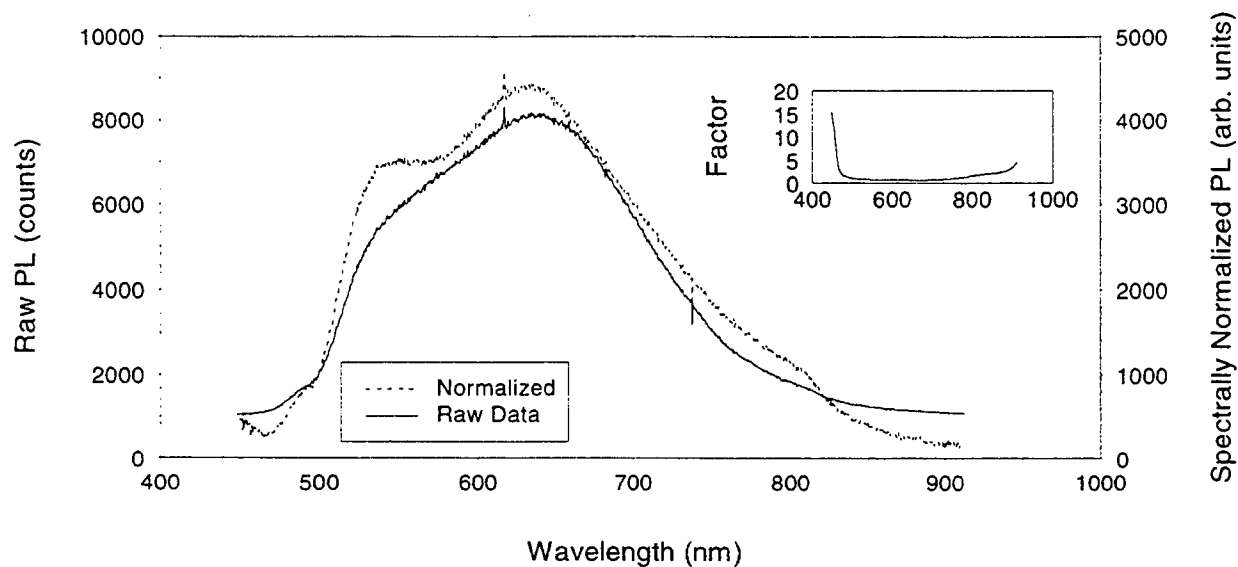## Room Temperature PL of As Received Spin Cast PBZT Film

Figure 3. PL of as received, spin cast, PBZT. The solid line represents the raw data, while the broken line is the result of spectral normalization. The inset shows the multiplicative normalization factor.
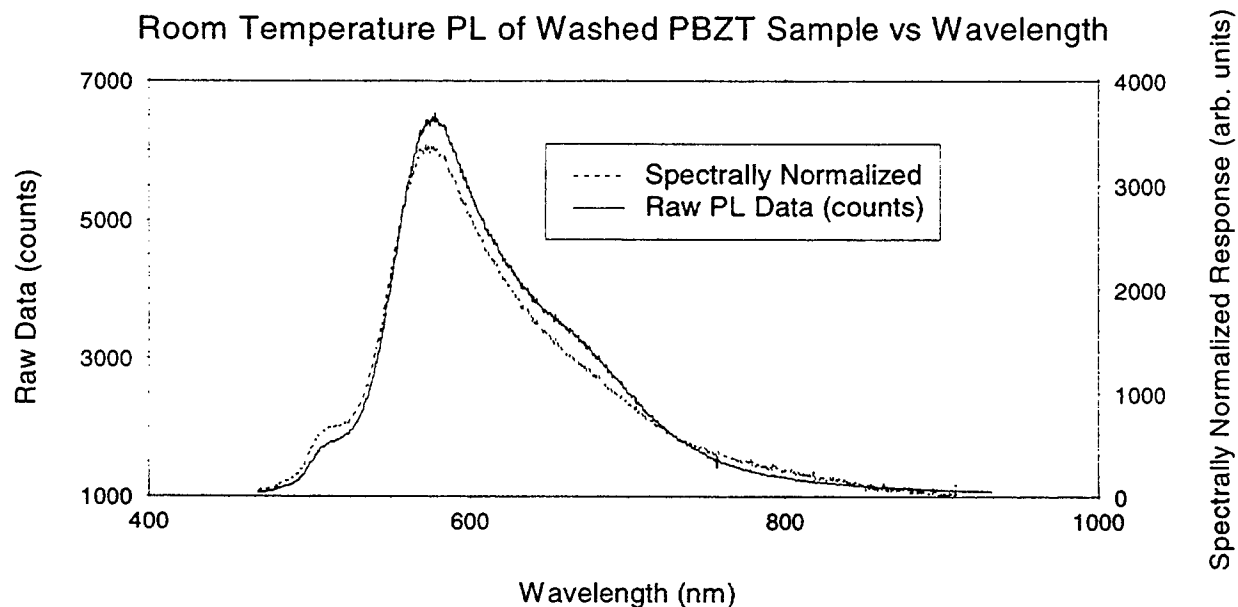
Room Temperature PL of Washed PBZT Sample vs Wavelength

Figure 4. Room temperature PL spectrum of washed PBZT sample. Solid curve is the raw data and the broken curve is the spectrally normalized data.

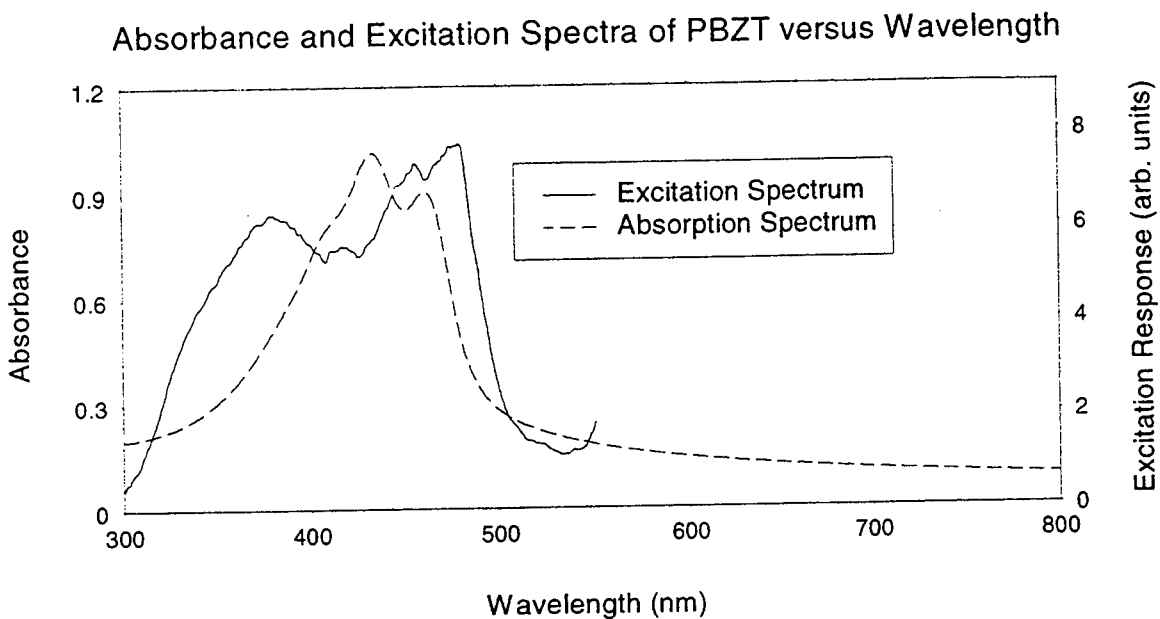## Absorbance and Excitation Spectra of PBZT versus Wavelength

Figure 5. Excitation and Absorption of PBZT. The excitation is for the washed sample, and the absorption is from a cast sample made especially for absorption studies.

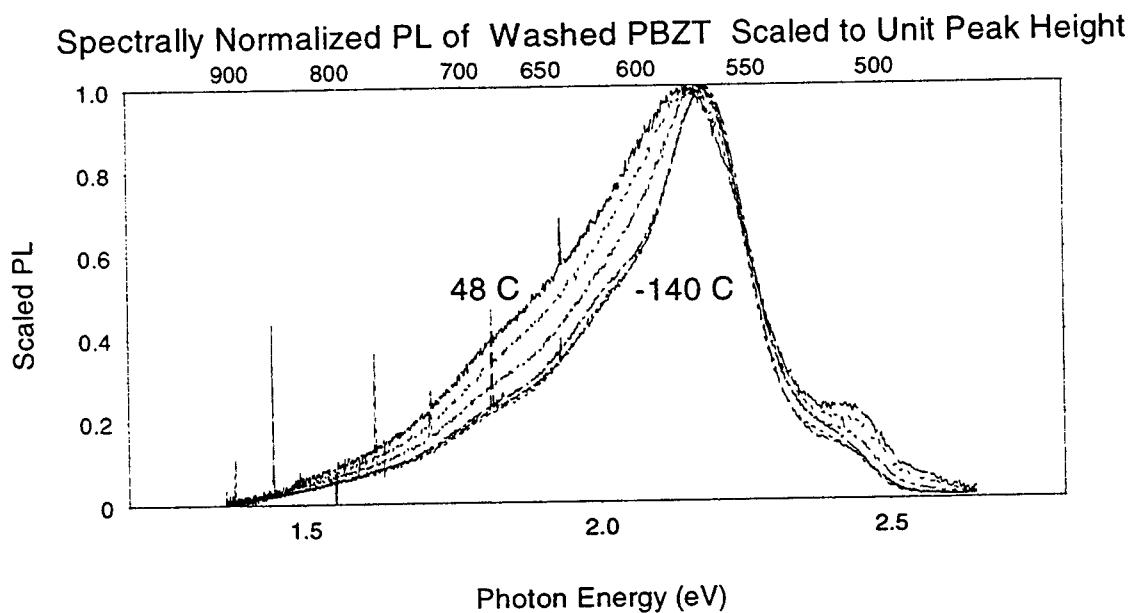## Spectrally Normalized PL of Washed PBZT Scaled to Unit Peak Height

Figure 6. Unity scaled spectrally normalized PL data for the washed PBZT sample. The increase in width relative to height at increased temperatures is obvious. The spikes are due to cosmic rays.
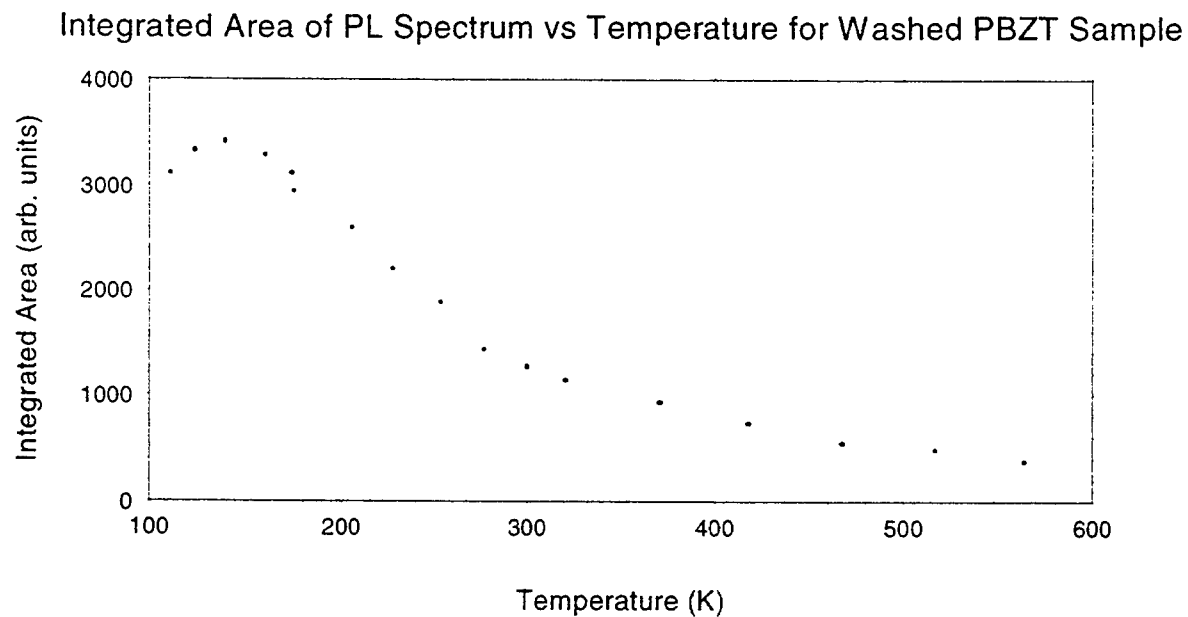
Integrated Area of PL Spectrum vs Temperature for Washed PBZT Sample

Figure 7. Integrated area vs. temperature of washed PBZT sample.

Integrated PL Intensity for Washed PBZT vs 1000/Temperature
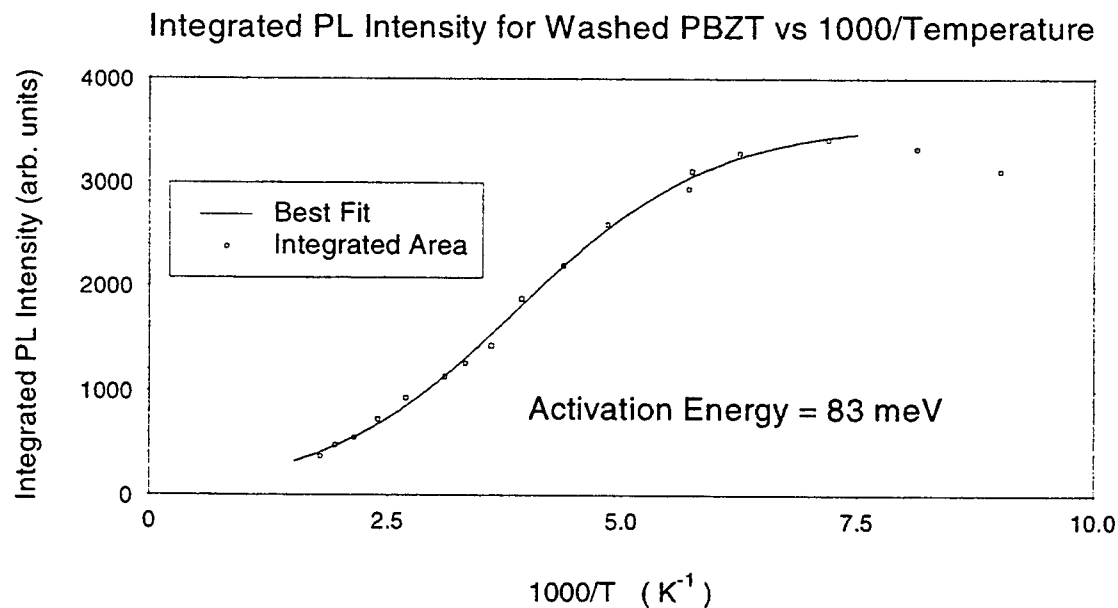

Activation Energy = 83 meV

Figure 8. Total Area vs. inverse temperature with fit. The solid line is the best fit function assuming thermally activated nonradiative recombination centers. The activation energy is 83 meV.

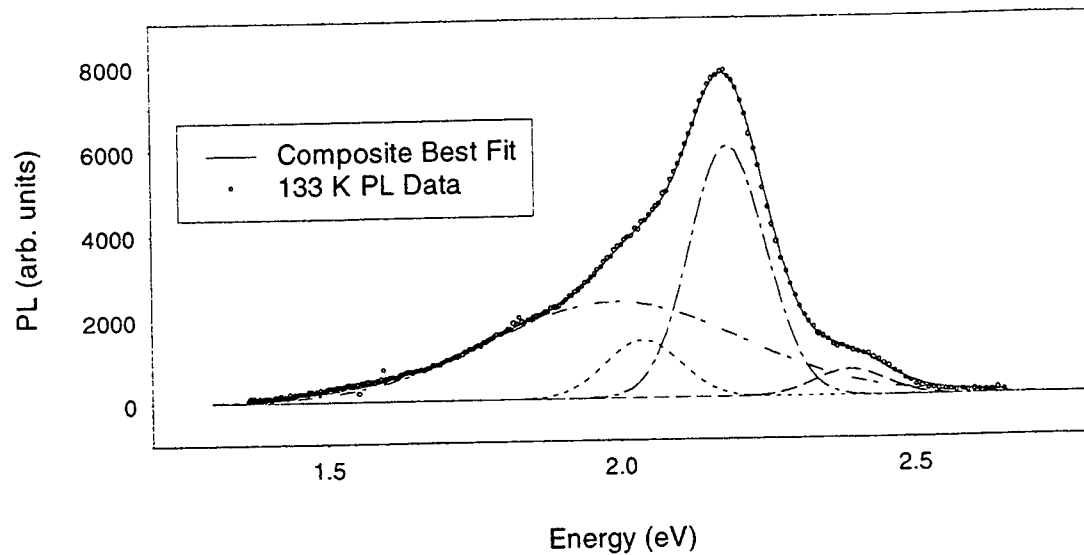## Gaussian Curves from Fit to 133 K PBZT PL Data

Figure 9. Results of fitting to 133 K PL data with multiple gaussians. The solid line is the composite curve, and the broken lines are the best-fit gaussians. Every fourth data point is marked by a symbol.

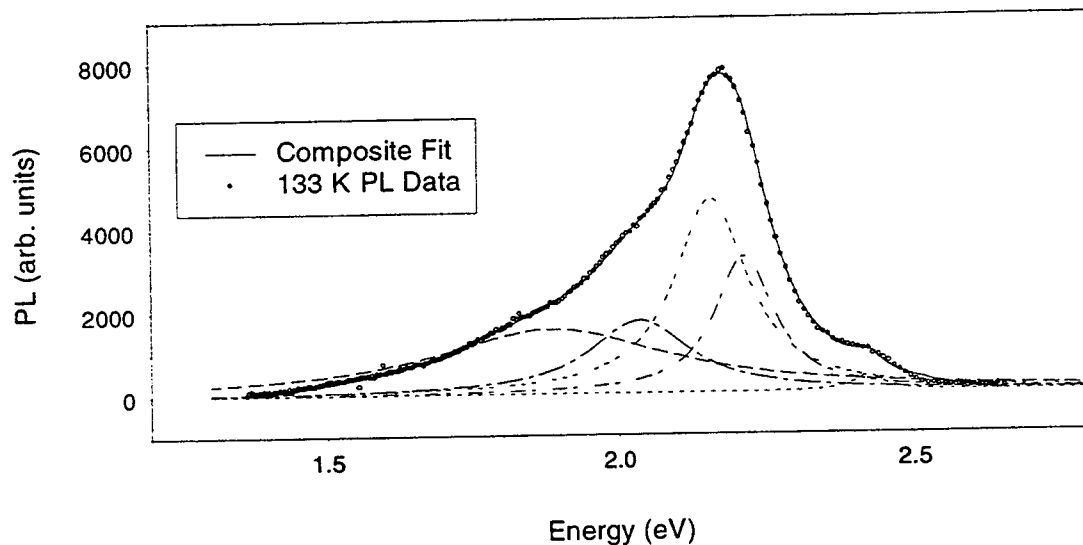## Lorentzian Curves from Fit to 133 K PBZT PL Data

Figure 10. Results of fitting to 133 K PL data with multiple lorentzians. The solid line is the composite curve, and the broken lines are the best-fit gaussians. Every fourth data point is marked by a symbol.

Best-Fit Parameters of Washed PBZT PL Temperature Study

		Gaussian # 1			Gaussian # 2		
Temp.	Offset	Amp.	Width	Energy	Amp.	Width	Energy
(k)	(counts)	(counts)	(eV)	(eV)	(counts)	(eV)	(eV)
111	64	1987	0.0680	2.0391	868	0.0625	2.3808
123	94	2166	0.0675	2.0372	1035	0.0624	2.3793
133	57	1376	0.0642	2.0392	617	0.0600	2.3929
139	113	2204	0.0688	2.0372	1124	0.0625	2.3784
160	104	2004	0.0685	2.0367	1109	0.0647	2.3774
174	91	1884	0.0719	2.0393	1006	0.0649	2.3778
175	80	1719	0.0714	2.0398	941	0.0676	2.3749
206	161	1461	0.0722	2.0392	849	0.0637	2.3911
228	57	1113	0.0759	2.0371	602	0.0598	2.4070
254	42	776	0.0785	2.0447	424	0.0539	2.4204
277	17	573	0.0832	2.0610	303	0.0467	2.4339
300	7	466	0.0808	2.0613	264	0.0434	2.4403
321	-3	435	0.0881	2.0691	238	0.0434	2.4423
371	-15	365	0.0793	2.0712	194	0.0433	2.4456
418	-33	308	0.0743	2.0806	162	0.0453	2.4487
468	-205	246	0.1464	1.9629	149	0.0443	2.4324
517	-108	479	0.1586	1.9684	351	0.0549	2.4278
564	-116	239	0.1259	1.9034	361	0.0582	2.4285

		Gaussian # 3			Gaussian # 4		
Temp.		Amp.	Width	Energy	Amp.	Width	Energy
(k)		(counts)	(eV)	(eV)	(counts)	(eV)	(eV)
111		7572	0.0629	2.1862	2518	0.2165	1.9848
123		7962	0.0638	2.1860	2664	0.2113	1.9801
133		5958	0.0651	2.1865	2286	0.2167	1.9927
139		7909	0.0647	2.1873	2736	0.2097	1.9778
160		7389	0.0656	2.1895	2710	0.2088	1.9790
174		6827	0.0656	2.1920	2563	0.2138	1.9813
175		6556	0.0652	2.1934	2454	0.2118	1.9809
206		5080	0.0714	2.1862	2203	0.1963	1.9680
228		4098	0.0727	2.1839	1965	0.2200	1.9837
254		2927	0.0742	2.1830	1911	0.2244	2.0066
277		1816	0.0736	2.1813	1597	0.2306	2.0207
300		1412	0.0748	2.1787	1476	0.2352	2.0216
321		1084	0.0722	2.1758	1377	0.2395	2.0252
371		729	0.0678	2.1720	1190	0.2457	2.0333
418		452	0.0615	2.1708	970	0.2556	2.0452
468		568	0.0885	2.1466	622	0.4202	2.0872
517		978	0.0961	2.1562	666	0.4269	2.0780
564		864	0.1046	2.1588	553	0.4696	2.0979

Table 1.   Best-fit results of fitting four gaussian lines to the washed PBZT
PL versus temperature data.

EVALUATION OF CONCRETE
BURST DAMAGE ALGORITHMS
IN EVA-3D

Joseph W. Tedesco
Gottlieh Professor
Department of Civil Engineering

Auburn University
Auburn, AL 36849

Final Report for:
Summer Faculty Research Program
Wright Laboratory, Armament Directorate

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Wright Laboratory

September 1995

EVALUATION OF CONCRETE
BURST DAMAGE ALGORITHMS
IN EVA-3D

Joseph W. Tedesco
Gottlieh Professor
Department of Civil Engineering
Auburn University

Abstract

EVA-3D is an effectiveness/vulnerability computer program which contains several algorithms for a prediction of damage sustained by hard target concrete structures subject to bursts in air, soil, and concrete. These algorithms were evaluated for their validity and accuracy. For the most part, these algorithms were not developed from sound engineering practices and extended available hard target vulnerability data beyond their limitations without implementation of a valid methodology. Future damage assessment algorithms should be based upon well established numerical and material models, and should be verified with existing hard target data.

EVALUATION OF CONCRETE
BURST DAMAGE ALGORITHMS
IN EVA-3D

Joseph W. Tedesco

Introduction

Effectiveness/Vulnerability Assessments in Three Dimensions (EVA-3D) is a computer program (Ref 1) which assesses the survivability/vulnerability of fixed hardened targets and the interaction of conventional weapons with those targets. It does this by modeling the attack or delivery conditions, the weapon trajectory, fuzing, and burst damage inflicted on the target in a Monte Carlo type of approach. The scope of this report is limited to concrete burst damage algorithms. The objective is to evaluate the validity and effectiveness of these algorithms.

EVA-3D contains several algorithms to predict the extent of damage due to bursts in air, in soil, and in concrete. Both collapse and partial damage algorithms are included. There are three generally recognized damage mechanisms for concrete components (i.e. external walls, interior wall, floors and ceiling): edge shear, localized shear (breach), and flexure (Ref 2). These damage mechanisms are illustrated in Figure 1.

For the edge shear mechanism, a complete severance or separation of the slab (or beam) occurs. For localized shear or breach, a complete severance or separation of a small volume of the structure from the slab (or beam) occurs. Flexural failure occurs when plastic hinges form at locations where the ultimate bending capacity is attained. Edge shear and localized shear are the predominant failure modes for close proximity bursts. The edge shear mechanism is tantamount to total failure or collapse. The localized shear (breaching) and flexural failure mechanisms are generally viewed as partial collapse and are capable of resisting load from subsequent bursts, albeit at a reduced capacity.

In EVA-3D structural damage or collapse can occur in three cases (Ref 3). The first is when a weapon detonates in the soil near enough to the target to breach a wall or panel. The second is similar to the first but can occur in either exterior air (outside the target structure) or interior air (inside the target room). The third is when a weapon bursts within a concrete component.

Soil Burst

A breach algorithm was developed from "first principles" to determine the shear failure of a concrete slab due to a detonation in adjacent soil (Ref 4). The soil breach algorithm was incorporated into the computer program BREACH (Ref 5). Input for the computer program includes soil conditions, slab properties, the weapon's explosive yield, and burst distance from the slab. The algorithms in BREACH calculate the free field environment due to the burst, transform the free field stresses into a load on a slab, and compare the load to the shear capacity of the slab. The free field peak stress is determined from the following equations:

$$V_p = \frac{21,225.7}{\sqrt{\rho_o}} (100)^n \left(\frac{r/3.28}{\left(W/2 \cdot 10^9 \right)^{1/3}} \right)^{-n} \qquad (1)$$

$$C_L = C_o + S \cdot V_p \qquad (2)$$

$$S = \frac{1.75}{1 + C_s/3280} \qquad (3)$$

$$\sigma_p = \rho_o \cdot C_L \cdot V_p \qquad (4)$$

where

V_p —   peak particle velocity (ft/s)
ρ_o —   soil density (lbs/ft³)
r —   range (ft)
W_e —   TNT explosive weight (lbs)
n —   attenuation factor
C_o —   wave speed (ft/s)
C_L —   loading wave speed (ft/s)
C_s —   seismic wave speed (ft/s)
σ_p —   peak stress (psi).

The free field stress time history is then calculated using an exponential decay given by

$$\sigma(t) = \sigma_p \exp\left(\frac{-(t - t_a)}{t_a} \right) \qquad (5)$$

in which t_a is the arrival time. The load on the slab is then determined by multiplying the free-field stress time history by a reflection factor that is a function of the incident angle of the stress wave as it impacts the concrete. The resulting load is a distribution of stresses over space and time. The impulse, I, imparted to the slab is calculated as the integral of stress applied to the loaded area over time.

To determine whether a breach in the slab occurs, a critical impulse is determined, as suggested by Ross (Ref 6), from the expression

$$I_{cr} = \frac{(2\sqrt{2})h}{3} \left[\left((1-q)\rho_c + q \cdot \rho_r \right)\left((1-q)\sigma_{cd} + q \cdot \sigma_{rd} \right) \right]^{1/2} \qquad (6)$$

where

I_{cr} —   critical impulse (psi - sec)

$$h \quad = \quad \text{thickness (in)}$$

$q \quad = \quad$ percent steel reinforcement (decimal value)

$R_c \quad = \quad$ concrete density (lb/in^3)

$\rho_c \quad = \quad$ steel density (lb/in^3)

$\sigma_{cd} \quad = \quad 3$ to $11 \sqrt{f_c'}$; $f_c' =$ concrete compressive strength

$\sigma_{rd} \quad = \quad$ tensile strength of steel reinforcing

When I exceeds I_{cr} then shear failure at the edges is assumed to occur. For localized shear (breaching), the solution algorithm is an iterative process that varies the radius of the breach hole until the loading impulse just exceeds the critical impulse of the slab as discussed in References 2 and 6. The program determines whether the slab was breached or not, and if it was breached, the size of the breach hole.

Breach equations that represent "cases of interest" (Ref 4) were developed for EVA-3D using the BREACH computer program. Bounding conditions were chosen that provided "worst case" (Ref 4) conditions. Those bounding conditions are presented in Table 1.

Table 1. Bounding Conditions for Soil Breach Equations

	Condition 1	Condition 2
Concrete Compressive Strength (psi)	3000	5000
Percent Steel Reinforcement	.5%	1%
Steel Yield Stress (psi)	40,000	60,000

The resulting hole size (scaled breach radius) versus scaled range values are illustrated in Figure 2 and Figure 3 for Conditions 1 and 2, respectively. The results are presented for slab thicknesses of 1 ft., 2 ft, and 3 ft. Review of the results for the two conditions indicated that the mean (for each concrete thickness) was "within the scatter of the results" (Ref 4). Thus, a single mean and standard deviation were developed and input into the EVA-3D computer code. A plot of the soil breach equations and the standard deviation (3σ) are presented in Figure 4. The curves represent the largest circular plug that can withstand the impulse in one quarter of the slab's natural period. It is alleged (Ref 7) that the 1 ft and 2 ft breach curves compare well to NDRC test data (Ref 8). Furthermore, it is also stated in Reference 7 that the shear mechanisms do not apply to slabs thicker than 3 ft.

Partially Damaged Concrete (Soilburst)

In many attack scenarios against an underground hard target, the penetrating weapon detonates in soil near a roof or wall. The concrete slab may be breached or damaged, providing less resistance to a subsequent weapon passing through the slab. In an attempt to address the failure mechanisms exhibited by thicker slabs (greater an 3 ft), computer simulations were performed (Ref 7) to create a more thorough concrete damage algorithm. Failure mechanisms and reduced

concrete capacity were determined from a series of 2-D, axisymmetric finite element calculations. In these calculations, an uncased charge was detonated in a soil medium at a prescribed standoff distance from a supported concrete slab. The following matrix of parameters was employed in the study:

Explosive Weight (lbs)	100, 1000
Scaled Stand off Distance (ft/$W^{1/3}$)	0.5, 0.75, 1.0, 1.25, 1.5
Steel Reinforcement (percent)	0.0 , 0.5, 1.0
Slab Thickness (in)	39.4, 78.7, 118.1

The extent of concrete damage was based upon axisymmetric SAMSON 2 (Ref 9) finite element calculations. The SAMSON2 computer program is a two-dimensional finite element code designed primarily for dynamic analysis of plane and axisymmetric solids. The main features of the code include nonlinear material modeling, explicit time integration, large displacement sliding/separating interfaces, and multiple time step integration.

Four material models were used in the SAMSON2 analyses: explosive, concrete, soil, and steel. The explosive was modeled by a spherical cavity with a pressure boundary at the soil-cavity interface. The pressure at the interface was determined using the Jones-Wilkens-Lee (JWL) equations of state parameters (Ref 10). The explosive type modeled in the calculations was H-6, a composite high explosive (HE) with non-ideal behavior.

The concrete and sand were modeled using Applied Research Associates' (ARA) 3-Invariant, Applied Engineering Cap Model (Ref 11, 12). The concrete used in all calculations had an unconfined compressive strength of 5000 psi and a density of 145 pcf. The soil in the calculations was described with a Fork Polk sand model. This type of sand is stiff and dry, having an internal friction angle of 31° and a density of 116 pcf. The reinforcing steel was modeled as an elastic-plastic material with a yield stress of 67,000 psi.

In the SAMSON 2 calculations, an explosive cavity was modeled in the soil medium centered above the concrete slab at a prescribed stand-off distance. The slab's horizontal bottom surface extended 116 in. from the centerline (axis of symmetry), and was supported at its end with a 40 inch thick concrete wall. The wall extended 8 ft. below the horizontal slab to minimize any wave reflection effects from the mesh's lower boundary. The soil elements also extended at least 8 ft. below the horizontal slab (outside the wall). Radial and circumferential bar elements (axisymmetric model) were placed in the concrete to model the reinforcing bars. A typical mesh used in the SAMSON2 calculations is presented in Figure 5.

The calculations were run for 15 to 18 ms. after detonation. Most of the damage was expected to occur during this time frame. The slab was considered to have failed when large vertical displacements occurred (Ref 7). A large vertical displacement was considered to be indicative of a breach. The size of the breach holes were determined from the extent of strain softening in the concrete.

The results of the finite element calculations were compared with the NDRC soil breach curves (Refs 8, 13). It was claimed (Ref 7) that the finite element calculations followed the NDRC breach curve "reasonably well". This comparison is illustrated in Figure 6.

The extended algorithm for concrete damage due to detonation in soil was developed for EVA-3D based on the results of the SAMSON2 calculations and on the NDRC soil burst curves. The damage and the corresponding reduction of penetration resistance was determined. EVA-3D determines if the concrete component fails based on the NDRC breach curve. If the NDRC breach curve does not indicate failure, EVA-3D uses the NDRC scabbing curves to calculate any reduced capacity in the concrete.

If the NDRC breach curve indicated failure, the concrete component is assumed to have failed in flexure at its support locations, or by punching shear. If the slab thickness is greater than 3 ft., the punching shear mechanism is "not probable" (Ref 7), so the slab is assumed to have failed at its supported edges (in flexure). Any residual concrete left is assumed to provide negligible resistance to a subsequent penetrating weapon. For slabs having a thickness of 3 ft or less, punching shear is deemed a possibility. In such cases, the ARA soil breach curves (Figure 4) are checked. If the ARA soil breach curves provide a zero breach radius, the concrete is assumed to have failed at the edges. The concrete component is removed at its supported edges and replaced by air. Should the ARA soil breach curves return a non-zero breach radius, the component fails due to punching shear. A cylindrical hole is placed in the concrete component, but the remainder of the component remains intact.

If the concrete does not fail (no breach), it may have a zone of reduced capacity concrete. Within this zone the concrete is assumed to be unconfined, providing less resistance to any subsequent penetrating weapon. The damaged zone is assumed to be spherical in shape with its center outside the concrete component. The burst point and the center of the damage sphere are an equal distance from the concrete component's back face (Figure 7). The damaged concrete depth into the component is found by plotting the point defined by the scaled burst range to the concrete component and the component's scaled thickness in Figure 7. Points which fall between curves are linearly interpolated to determine the spall zone.

Air Burst

The peak free-field pressure associated with the detonation of a spherical loading charge is given by the classical Hopkinson Scaling law (Ref 2):

$$P = k\left(\frac{R}{W^{1/3}}\right)^{-n} = k \cdot \lambda^{-n} \tag{7}$$

where

P	=	pressure (psi)
k	=	intercept of $\lambda = 1$ on $\log P$ vs $\log \lambda$ curve
$-n$	=	slope of curve
R	=	stand off distance (ft)

$$W \quad = \quad \text{charge weight (lbs)}$$

In EVA-3D $k = 10^{3.051}$ and $n = 2.197$. The coefficients are supposedly based upon NDRC test results for cased and uncased charges (Ref 14). This pressure formula is applicable to detonations in both "exterior air" (outside the target structure), and "interior air" (within a target room), with some modifications.

A detonation within a structure produces both shock and air pressure loads. The airblast shock loads include the initial incident wave and the reflections from adjacent surfaces. The enclosed structure restricts the venting of gaseous ventilation products resulting in relatively long duration gas pressure loads. The rise and decay times of the gas pressure loads can be very long relative to those of the shock pressure. The long duration gas pressure is a result of restricted venting and dominates the long term impulse. The explosive weight and location, structure geometry and volume, vent area and frangible surfaces all affect the magnitude and duration of these loads (Ref 15).

In EVA-3D, the explosive weight (or yield) used to calculate the peak pressure (Figure 7) is enhanced if the burst point is sufficiently close (within $1.5W^{1/3}$) to a reflective surface such as a wall, ceiling or floor. The resulting enhanced yield is used in all subsequent calculations of peak pressure. The enhancement yield (Ref 1) is given by

$$W_e = W \cdot 2^n \tag{8}$$

where n is the number of reflective surfaces ($n \leq 3$). If there are no reflective surfaces, the height of the room is checked for possible reflective effects. If the room height is less than $4W^{1/3}$ then the explosive weight is doubled. If the room height is between $4W^{1/3}$, and $5W^{1/3}$, then the explosive weight is enhanced by a factor of 1.6. This yield enhancement procedure developed for internal air blast is detailed in Reference 3. The yield enhancement factors were determined from the results of an unreferenced "previous parametric study" (Ref 3) using the computer codes BLAM (Ref 16) and BLASTIN (Ref 17).

Blast in exterior air environments is treated identically to interior airblast with the exception that roof and ceiling effects are not considered. The explosive weight is again enhanced if the burst location is within $1.5W^{1/3}$ of a reflective surface (Ref 1). No documentation pertaining to the enhancement methodology for exterior air blast could be retrieved.

Air Breach

The original air breach formula was developed as a study on target vulnerability (Ref 8). It was based upon data derived from tests, both model and full scale, or rectangular panels with face dimensions from 3 to 25 times the panel thickness. Charges used in the tests ranged from approximately 1 1/4 oz to 1000 lbs. Test panels were supported along all four edges, and results showed no appreciable difference between simply supported and fixed edges.

Tests involved various degrees of reinforcement and different explosives, but all data were reduced to a basis of 1/4% steel reinforcement by volume and bombs filed with TNT. Damage was categorized as slight, moderate, heavy, or breaching. A description of these damage categories is provided in Table 2. The degree of damage and ratio of central deflection to span length correlated fairly well for slight, moderate, and heavy damage (Ref 8).

The original air breach equation is given as

$$\frac{t}{W^{1/3}} = \frac{1}{\frac{1}{A} \cdot \frac{r}{W^{1/3}} + B} \tag{9}$$

where

$$\frac{t}{W^{1/3}} = \text{scaled thickness}$$

$$\frac{r}{W^{1/3}} = \text{scaled range}$$

and A and B are constants given in Table 3. The formula is valid for approximately 1/4% reinforcing steel by volume. It is suggested (Ref 8) that the formula may be extended to 1/2% reinforcing by multiplying the thickness by 0.9 for breaching and by 0.7 for all other degrees of damage. It is noteworthy that no information pertaining to the material properties of the concrete and the steel used in these tests was furnished. Graphs of the NDRC air breach formulae for breaching, heavy scabbing, and moderate scabbing are presented in Figure 8.

There are several important limitations of the NDRC Air Breach Formula. First, the equation is limited to structures having the same concrete and steel material properties of the test structures. Unfortunately, those material properties are not disclosed in any of the NDRC test data. Furthermore, the equation is only applicable to structures having either 1/4% or 1/2% reinforcement ratio. The equation is valid only for a TNT explosive, vertical weapon orientation and exterior air bursts. Finally, the size of the breach (breach radius) can not be predicted from the equation.

An attempt was made to extend the NDRC Air Breach Formula (Ref 5) to include a wide range of material properties and reinforcement ratios. An attempt was also made to determine the size of the breach, or breach radius. In this study (Ref 8) the concrete compressive strength and the steel yield strength for the NDRC test structures were "assumed" to be 3000 psi and 40,000 psi, respectively.

Table 2. Damage Criteria for Air Breach Formula (Ref 8)

Description of Damage	Type of Damage	Midspan Deflection / Span Length (in/ft)
Slight	Slight cracking and bending.	0.1
Moderate	Light punching and cracking with possibly some spalling.	0.5
Heavy	Heavy punching, shattering, or possible perforation.	1.2
	Perforation with extensive scabbing. Bars may be bent or bulged.	---

Table 3. Coefficients for Air Breach Formula (Ref 8)

Damage	Coefficients	
	A	*B*
Slight	0.32	0.20
Moderate	0.25	0.57
Heavy	0.24	1.50
Breaching	0.18	2.10

The general approach taken in the development of the air breach algorithm for EVA-3D was to (1) perform a series of bounding calculations which assess concrete breach due to either shear or flexural failure, (2) develop appropriate breaching equations for these bounds, and (3) perform a statistical analysis of the results leading to a statistical equation denoting the likelihood of breach hole size.

The details of the analyses conducted in the development of the air breach equation are presented in Reference 8. A parametric study was performed using the computer programs BREACH (Ref 8, Appendix A), TBREACH (Ref 8, Appendix B), and FLEXURE (Ref 8, Appendix C). The variables included two slab widths (10 ft and 50 ft), five explosive weights ranging from 50 lbs to 2000 lbs, two steel strengths (40 ksi and 60 ksi), three concrete strengths (3000 psi, 4000 psi, and 5000 psi), four slab thicknesses ranging from 0.5 ft to 3.0 ft and two reinforcement ratios (.5% and 1.0%).

Using the computer program BREACH, which incorporates the Speicher-Brode pressure/time history formulation, breach curves of shear failure described in terms of scaled breach radius were generated. The algorithm was essentially the same as that described for the soil breach developed by Ross (Ref 2). For all combination of data analyzed, smooth circular curves were generated. Each curve, however, had a different radius. Therefore, in an attempt to develop a single curve representative of breach radius, the generated family of curves were collapsed into one curve of one radius.

In an effort to determine "the correct multiplier" to collapse the curves, another series of analyses were conducted using the computer program TBREACH. These analyses, in turn, spawned yet another family of curves describing breach as a function of scaled thickness versus scaled range. The resulting curves were plotted against the NDRC air breach curve where "some similarity" was noted.

The curves generated by TBREACH were then collapsed into a single curve for which a best fit could be found, and for which a "near perfect" correlation with the NDRC Air Breach Curves was proclaimed, Figure 9 (Ref 8). Having collapsed the breach curves on a graph of scaled thickness versus scaled range, they still could not be successively collapsed on a graph of scaled breach radius to scaled range. However, this result was proclaimed to have been achieved (Ref 8) by multiplying both the breach radius and the range by the factor

$$\frac{t}{W^{1/3}} + 0.5$$

This manipulation allegedly extended the NDRC Air Breach Formula to a larger database, and could also describe breach size (Figure 10).

This extended breaching curve was representative of shear type breaches only. Therefore, it was deemed necessary to develop a curve which represented damage induced by both shear and flexural failure. To this end, curves representing failure due to flexure only were generated by conducting a parametric study (using the same variables previously cited) by implementation of the FLEXURE computer program. The flexural capacity of a concrete structural member was defined (Ref 4) as a function of the pressure acting on the slab. The maximum pressure that can be sustained by the slab as a function of the flexural capacity was given by (Ref 4) the following expressions:

$$PM_c = q_y \left(\frac{h}{\pi \cdot t_d} \cdot \sqrt{2\mu - 1} + 2\mu \left(\frac{2\mu - 1}{1 + \frac{2h}{\pi \cdot t_d}} \right) \right) \tag{10}$$

where

PM_c	=	average pressure capacity (psi)
t_d	=	effective positive phase loading (seconds)
μ	=	$\dfrac{\text{failure displacement}}{\text{yield displacement}}$
h	=	slab thickness (in)
q_y	=	flexural capacity (psi), which is calculated as follows:

$$q_y = 7.2 \left(p_c + p_e \right) \cdot f_{dy} \cdot \frac{b}{a} \cdot \left(\frac{d}{L} \right)^2 \tag{11}$$

where

p_c, p_e = tensile steel ratios at the center and ends
f_{dy} = steel dynamic yield strength (psi)
b = beam width (ft)
a = width of contributory load area (ft)
d = depth to the center of the steel (in)
L = length (ft)

For a one-way slab, $b = a$. Therefore, $b/a = 1$ and $d = h - 2$, assuming a 2 inch cover.

With these adjustments, the average capacity equation simplified to:

$$q_y = 14.4 \cdot p_c \cdot f_{dy} \left(\frac{h-2}{L}\right)^2 \qquad (12)$$

assuming p_c and p_e are equal. The average loading pressure is then calculated as

$$PM_L = \tfrac{1}{3}(P_2 - P_1) + P_{avg} \qquad (13)$$

where

PM_L = average pressure load (psi)
P_1 = pressure at the corners (psi)
P_2 = pressure at the center (psi)
P_{avg} = P_1 since the slab is assumed to be symmetric

If PM_L is greater than PM_c, then flexural failure is assumed to have occurred.

Inspection of the resulting flexural failure curves indicated that only the "middle section" of the curves were significant. The "left" edge of this middle section was labeled the "minimum", and the far right edge the "maximum". A mean was estimated within each middle section. A plot of the minimum, maximum, and mean of each graph versus concrete strength indicated that they formed straight lines. These points were subsequently used (although the details of the methodology are somewhat ambiguous) to form a family of curves representing breach failure by both flexure and shear.

Thus, the revised breaching algorithm for combined shear-flexural failure is given (Ref 4) as

$$B_r^2 = \left(\frac{C \cdot W_e^{1/3}}{S}\right)^2 - R^2 \qquad (14)$$

where

C = $0.46 \pm 0.11 (3\sigma)$

$$S = \frac{t/W_e^{1/3}}{0.18} + 0.5$$

B_r = breach radius (ft)
R = range from detonation (ft)
W_e = equivalent explosive weight (lbs)
t = concrete wall thickness (ft)

The extended air breach algorithm is illustrated in Figure 11.

Panel Collapse

In target structures where the walls are assumed to be constructed of prefabricated concrete panels supported by beams and columns, the panel collapse mechanism is evaluated. This method addresses the edge shear mechanism and uses loads calculated as a function of the maximum peak pressure at each of the corners. The load on the panel is calculated as (Ref 1, 3, 18)

$$P_L = \left[\tfrac{1}{3}\left(P_{peak} - P_{avg}\right) + P_{avg}\right] \cdot A \tag{15}$$

where

P_L = load on panel (lb)
P_{peak} = peak pressure on slab (psi)
P_{avg} = average pressure at corners (psi)
A = area of slab (in^2)

In Reference 3 it is stated that Eq (15) has been evaluated for several distributed pressure cases and was found to produce total forces that are accurate to within 5%. However, no details pertaining to the nature of the "evaluation" were presented nor was the accuracy benchmark defined.

A total shear force capacity (or fragility), upper bound and lower bound, is calculated from the expression (Ref 1, 3, 18)

$$P_c = P_r \cdot T \cdot C \cdot \sqrt{f_c'} \tag{16}$$

where

P_c = panel shear capacity
P_r = perimeter of panel
T = panel thickness
C = constant (3 for lower bound, 11 for upper bound)
f_c' = concrete compressive strength

If P_L is greater than or equal to P_c (upper bound), then full panel collapse is assumed . If P_L is less than P_c (lower bound), then no collapse is assumed to occur. However, if P_L is between P_c

(upper bound) and Pc (lower bound), then partial collapse mechanisms in the form of localized shear and flexure are investigated.

Conclusions

Hard target vulnerability assessments must, of necessity, rely heavily on vulnerability data generated during the 1940's. A common concern shared by those who perform such assessments is that the data upon which these assessments are based are very old and hence suspect, not relevant or at least only partially relevant (Ref 13).

There are few basic problems associated with the terminal ballistics of hard target attack that have been completely solved. Thus, it is necessary to rely on empirical data and equations for such important estimates as concrete scabbing and breaching, and effects of internal detonations in concrete structures. Much of the data and analytical relationships currently in use were largely developed during the 1940's. If these data are used within their limitations, they are useful and valid. However, it is a well established technical principal that one should not extend an empirically based estimator beyond the range of its data. For example, the results of concrete scabbing/breaching tests conducted against lightly reinforced 3000 psi concrete structures (i.e., NDRC data) cannot be applied without reservation to estimating scabbing/breaching of heavily reinforced 10,000 psi concrete structures (Ref 13).

Most of the concrete damage algorithms incorporated into the EVA-3D computer program have unfortunately ascribed to this flawed practice. Not only have the limited data been extended well beyond their limitations, but the methodology employed to this end, in most instances, was fundamentally unsound. For example, the concrete breaching algorithm's, for both soil breach and air breach, were based upon the results of parametric studies which were quite limited in scope. Moreover, the analytical procedures used in these studies were for the most part not validated or verified.

A glaring example of this violation of a well founded scientific study was the extension of the soil breach algorithm to slabs thicker than 3 ft. In this study (Ref 7), a series of nonlinear dynamic finite element calculations were conducted. The finite element code employed in the study was SAMSON2 (Ref 9). SAMSON2 is an "explicit" code, that is, the time integration scheme is explicit in nature. Such codes are very effective for wave propagation analyses, in which the nonlinearities are relatively mild. In an explicit scheme, equilibrium iterations are not performed. Therefore, in cases where the nonlinearity becomes extensive, solution convergence cannot be ascertained (Ref 19). Not surprisingly, most of the example problems presented in the SAMSON2 User's Manual are linear wave propagation analyses. To compound the inaccuracies inherent of the study conducted in Reference 7, the material model used for the concrete and sand was not validated. No documentation verifying the accuracy of the ARA 3-Invariant, Applied Engineering Cap Model (Ref 11) could be found.

Other instances of unvalidated algorithms include the maximum sustainable pressure designating the flexural capacity of a slab (Eq 10), the calculated maxiumum load on a panel given by Eq. (15), and panel fragility predicted by Eq (16). Furthermore, no documentation relevant to bursts

within concrete components could be located. Therefore, the algorithms pertaining to this aspect of vulnerability assessment were not addressed.

Finally, all the algorithms pertaining to concrete damage developed for EVA-3D were based upon experimental data and analytical/numerical calculations for flat concrete slabs. However, in EVA-3D the option for modeling cylindrical concrete targets is available. The application of concrete damage algorithms developed for flat slabs to the analysis of cylinder or arch-like structures is reprehensible.

Recommendations

The data base for hard-target vulnerability assessments is indeed old. It is generally of excellent quality, but it does not always provide adequate information for modern hard target vulnerability assessments (Ref 13). Efforts to generate vulnerability data related to hard target attack have been pursued sporadically and at low levels since 1945. Unless the effort devoted to the generation of new vulnerability data is substantially increased, hard target vulnerability assessments will have to rely on old vulnerability data (Ref 13).

In lieu of new, extensive, large scale studies of hard target vulnerability (comparable to those conducted in the 1940's), hard target vulnerability assessment must be accomplished through properly implemented comprehensive analytical/numerical studies. These studies should be conducted through the employment of validated computer codes and material models, and used in conjunction with well established scientific fundamentals. The existing hard target vulnerability data should be used to verify and validate the accuracy of newly developed hard target vulnerability assessment algorithms.

A specific recommendation for the immediate improvement of the blast damage algorithms in EVA-3D would be to incorporate the computer program REICON (Ref 2) in the main module. However, the REICON program should be modified to accommodate air bursts and to allow for bursts at locations other than at the midpoint of the structure before it is implemented into the EVA-3D module.

Finally, it is recommended that a Verification Manual be developed for the EVA-3D computer program. This manual would contain relatively simple example problems in which results obtained from the various algorithms in EVA-3D can be compared with existing data or solutions to validate their accuracy.

References

1. Maestas, F.A., Young, L.A., Streit, B.K., and Peterson, K.J., <u>Effectiveness/Vulnerability Assessment in Three Dimensions (EVA-3D), User's Manual</u>, Version 4.1, Applied Research Associates, July 1995.

2. Ross, C.A., Sierakowski, R.L., and Schauble, C.C., <u>Concrete Breaching Analysis</u>, AFATL-TR-81-105, Air Force Armament Laboratory, December 1981.

3. Maestas, F. A. And Galloway, J.C., <u>Development of Target Models, Volume I: Methodology of EVA-3D</u>, AFATL-TR-89-13, Air Force Armament Laboratory, August 1989.

4. Hacker, W. L. , Maestas, F. A., and Galloway, J. C., Survivability Assessment Methodology (SAM) - Volume 1, ESL-TR-90-48, Air Force Engineering and Services Center, April 1991.

5. Young, L. A. And Maestas, F. A., Damage States: Extending the Air Breach Approach, Applied Research Associates, June 1988.

6. Ross, L. A. And Rosengren, P. L., "Expedient Nonlinear Dynamic Analysis of Reinforced Concrete Structures", <u>Proceedings of the Second Symposium on the Interaction of Non-Nuclear Munitions with Structures</u>, Panama City, FL, 1985, pp. 45-51.

7. Maesta, F. A., Cilke, R. W., and Hacker, W. L., <u>Cumulative Damage Methodology Development, Phase III, Final Report</u>, Aeronautical Systems Center, April 1993.

8. Stipe, J. G., <u>Summary Technical Report of the National Defense Research Committee, Vol I: Effect of Impact and Explosion</u>. "Weapon Data: Damage of Reinforced Concrete Wall Panels", Office of Research and Development, Washington, D.C., 1946.

9. Schreyer, H. L. Richards, C. G., Bean, J. E., and Durka, G. R., <u>SAMSON2, A Nonlinear Two-Dimensional Structure Media Interaction Computer Code: User's Manual</u>, New Mexico Engineering Research Institute, University of New Mexico, AFWL-TN-82-018, September 1982.

10. Dobratz, B. M., <u>LLNL Explosives Handbook, Properties of Chemical Explosives and Explosive Simulates,</u> Lawrence Livermore National Laboratory, Livermore, CA, UCRL-529997, Section 8.3.1, March 1981.

11. Bingham, B. L., <u>Applied Engineering Cap Model with Three Invariants,</u> Internal Report, Applied Research Associates, Albuquerque, New Mexico (no date).

12. Keller, J. A., Lesko, J. M., and Bingham, B. L., <u>Experimental Results for Material Modeling and Breaching Dynamics of Thick Concrete Slabs</u>, WL-TR-93-7054, Wright Laboratory Armament Directorate, October 1992.

13. Wolfersberger, J. R., <u>Concrete-Breaching Data Base Search</u>, ESL-TR-84-37, Air Force Engineering and Services Center, August 1984.

14. Frew, K. C., EVA-3D Version 1.5: User's Manual, WL-TR-93-7055, Wright Laboratory, Armament Directorate, July 1993.

15. McMahon, G. W. et.al., "Air Blast", <u>Joint Services Manual for the Design of Hardened Structures of Conventional Weapons Effects</u>, Chapter 5, Defense Nuclear Agency, June 1995.

16. Cobb, M. B., Drake, J. L., and Britt, J. R., <u>BLAM User's Manual</u>, Applied Research Associates, June 1986.

17. Britt, J. R., Drake, J. L., Cobb, M. B., and Mobley, J. P., <u>BLASTIN User's Manual</u>, Report No. ARA-5986-2, Applied Research Associates, April 1986.

18. Hacker, W. D. And Tucker, R. K., <u>EVA-3D Version 2.0 User's Manual</u>, Applied Research Associates, April 1991.

19. Bathe, K. J., <u>Finite Element Procedures in Engineering Analysis</u>, Prentice Hall, Inc., 1982.

EDGE SHEAR

LOCALIZED SHEAR

FLEXURAL
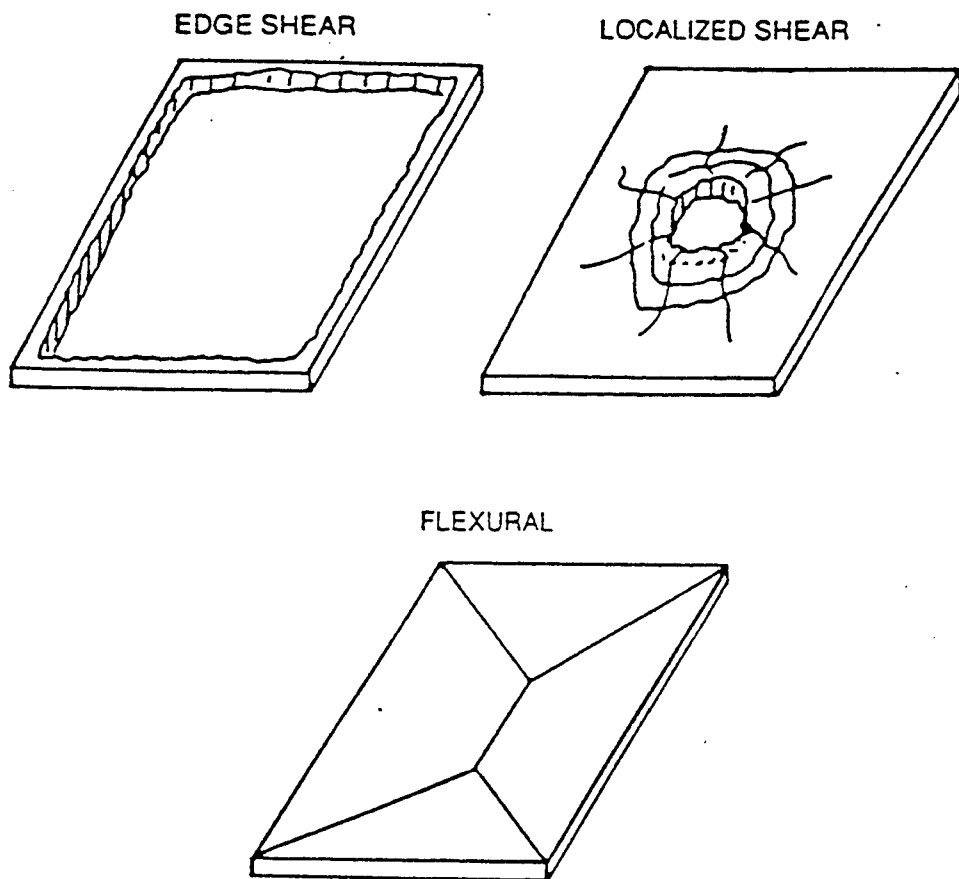
Fig. 1. Failure Mechanisms (Ref. 4)
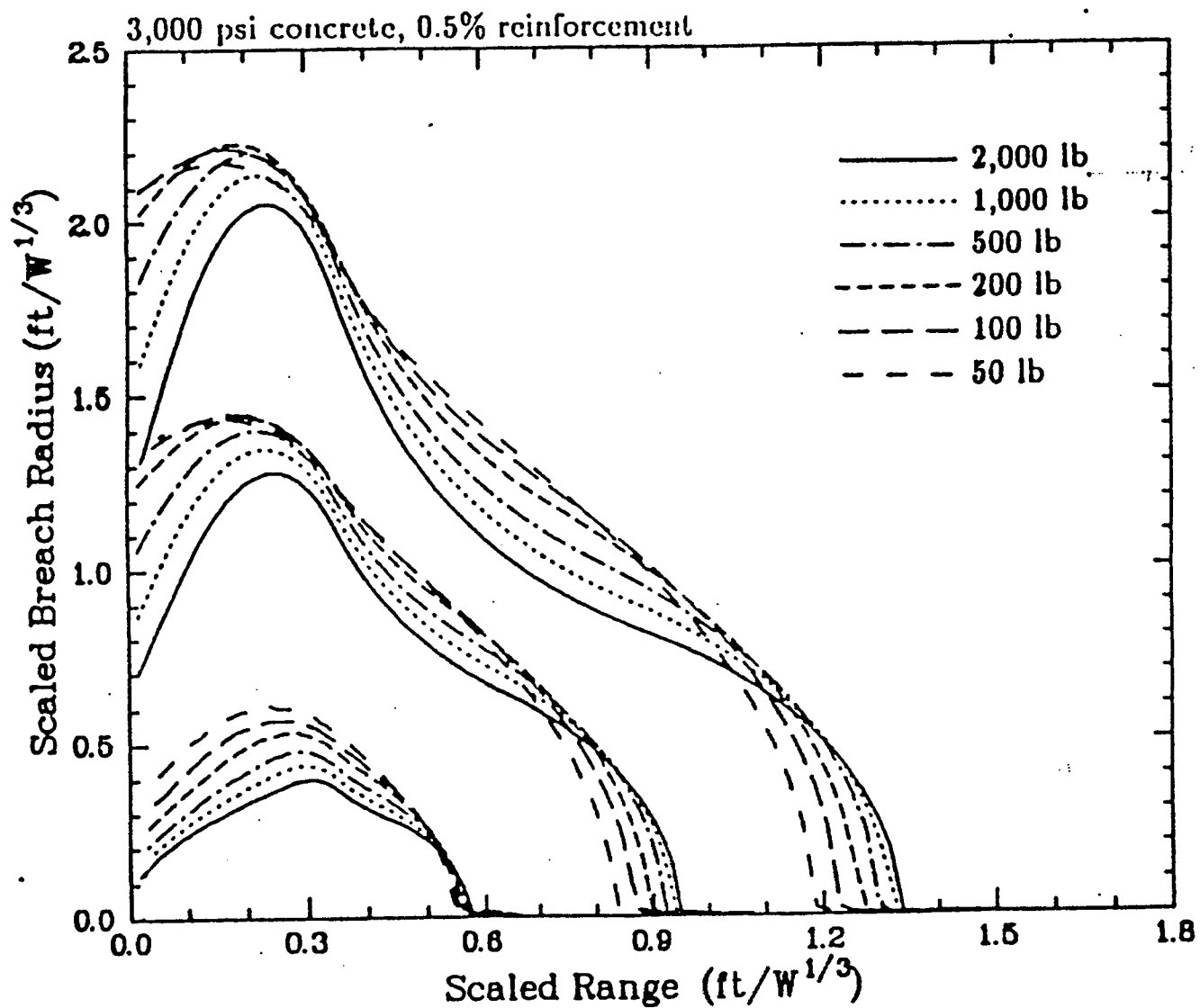
Fig. 2. Soil Breach Radius for Condition 1 (Ref. 4)

56-19
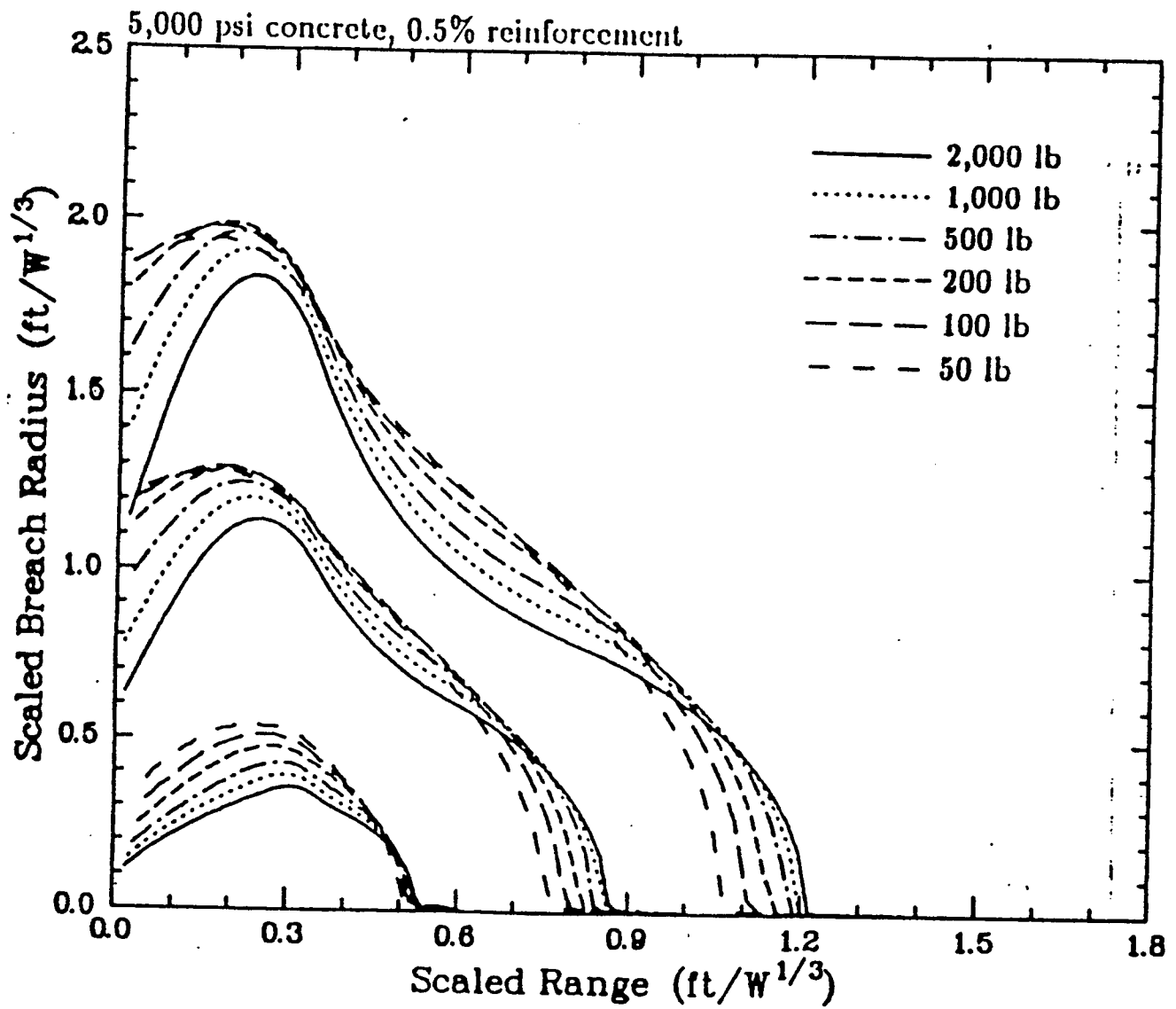
Fig. 3. Soil Breach Radius for Condition 2 (Ref. 4)

56-20

Structural Operational Semantics for a VHDL-93 subset

Krishnaprasad Thirunarayan
Associate Professor
Department of Computer Science and Engineering

Wright State University
3640, Col. Glenn Highway
Dayton, OH 45435.

Final Report for:
Summer Faculty Research Program
Wright Laboratory

September 1995

Structural Operational Semantics for a VHDL-93 subset

Krishnaprasad Thirunarayan
Associate Professor
Department of Computer Science and Engineering
Wright State University

September 25, 1995

Abstract

Goossens defined a structural operational semantics for a subset of VHDL-87 that includes delta delays, arbitrary wait statements, and (commutative) resolution functions. A monogenicity result was proved showing that the parallelism present in the VHDL is benign.

In this paper, we correct and extend Goossens work to include VHDL-93 features such as shared variables and postponed processes that change the underlying semantic model of VHDL. The monogenicity result does not hold in general when shared variables are introduced. However, we identify and characterize a class of *portable* programs for which the monogenicity result can be salvaged. Our specification can serve as a correctness criteria for the VHDL-93 simulator.

Structural Operational Semantics of a VHDL-93 subset

Krishnaprasad Thirunarayan

1 Introduction

VHDL is one of the most widely used hardware description languages. It has been designed to facilitate specification, documentation, communication and formal manipulation of hardware designs at various levels of abstraction [1]. The semantics of VHDL-93 is given in English prose in [7]. This kind of informal specification has potential to be interpreted differently by different people. The goal of developing a formal semantics of VHDL-93 is to provide a complete and unambiguous specification of the language. Adherence to this standard will contribute significantly to the sharing, portability and integration of various applications and computer-aided design tools; to the implementation of language processors; and for formal reasoning about VHDL programs.

There have been a number of proposals for a formal semantics of VHDL, almost all of them dealing with subsets of VHDL-87 [2, 3, 4, 8, 9]. For instance, van Tassel [8] presents an operational semantics of *Femto-VHDL*, a VHDL-87 subset that does not contain the *wait*-statement. Breuer *et al* [2] develop a denotational semantics and a proof theory for unit-delay VHDL. Goossens [4] defines a structural operational semantics [5] for a subset of VHDL-87 that includes local variables, signals (possibly resolved using commutative resolution functions), signal assignments (including zero-delay scheduling), arbitrary wait statements, and parallel composition of sequential programs. This work seems to be the most general covering almost all the fundamental behavioral constructs in a single VHDL-87 entity. A monogenicity result is also proved to show that the parallelism present in VHDL-87 is benign. Just recently, we also became aware of the work of Börger *et al* (Chapter 4, [3]) that provides a formal definition of VHDL-93 features using EA-machines. However, they do not seem to formally prove properties of their semantics or characterize when a VHDL-93 program containing shared variables is *erroneous*.

In this paper we build on Goossens work as reported in [4]. We define a structural operational semantics for a subset of VHDL-93 that includes features such as *shared variables* and *postponed processes* not present in VHDL-87. These VHDL-93 constructs fundamentally change the underlying semantic model of VHDL. In particular, the monogenicity property proved for the subset of VHDL-87 in [4] no longer holds in the presence of shared variables because of the non-deterministic and asynchronous nature of process executions. However, we characterize a class of *portable* VHDL-93 programs for which the monogenicity property can be salvaged. In other words, we identify a subset of VHDL-93 programs that can contain shared variables and yet admit a unique "meaning". The goal is to provide an approximate but formal interpretation of the following statement in Section 4.3.1.3 in the LRM [7].

A description is *erroneous* if it depends on whether or how an implementation sequentializes access to shared variables.

Our formalization can be viewed as a specification for the VHDL-93 simulators against which the correctness of an implementation can be verified. In course of this development we also explain and correct a few errors that have crept into the formal description of the VHDL-87 semantics given in [4].

The rest of this paper is organized as follows: Section 2 presents the abstract syntax of the VHDL-93 subset and Section 4 specifies its semantics. The primary emphasis is on the changes to the semantic model resulting from the introduction of shared variables and postponed processes. Section 5 presents conclusions and future work.

2 Syntax of VHDL-93 subset

The abstract syntax of the VHDL-93 subset is shown below:

- Syntactic Categories

pgm	$\in$	$Programs$	$proc$	$\in$	$Processes$
p	$\in$	$NonPostponedProcesses$	pp	$\in$	$PostponedProcesses$
ss	$\in$	$SequentialStatements(= SSt)$	e	$\in$	$Expressions(= Expr)$
s	$\in$	$Signals(= Sig)$	S	$\in$	$SetsOfSignals$
x	$\in$	$Variables(= Var)$	sx	$\in$	$SharedVariables(= SVar)$
v	$\in$	$Values(= Val)$			

- Definitions

$$
\begin{aligned}
pgm &::= \|_{i \in I} \ proc_i \\
proc_i &::= p_i \mid pp_i \\
p_i &::= \texttt{while } true \texttt{ do } ss_i \\
pp_i &::= \texttt{while } true \texttt{ do } ss_i \\
ss_i &::= \texttt{null} \mid x := e_i \mid sx_i := e_i \mid s <= e_i \texttt{ after } e_i \\
&\quad \mid ss_i \ ; \ ss_i \mid \texttt{wait on } S \texttt{ for } e_i \texttt{ until } e_i \\
&\quad \mid \texttt{while } e_i \texttt{ do } ss_i \mid \texttt{if } e_i \texttt{ then } ss_i \texttt{ else } ss_i \\
e_i &::= \texttt{null} \mid v \mid x \mid sx_i \mid s \\
&\quad \mid e_i \ bop \ e_i \mid uop \ e_i \mid s'\texttt{delayed}(e_i)
\end{aligned}
$$

A program in this VHDL-93 subset can be viewed as a behavioral VHDL-93 program or as a fully elaborated VHDL-93 program [7]. It is basically a collection of processes communicating with each other through signals and shared variables. || is the parallel composition operator and I is an arbitrary, finite index set for the processes. In order to characterize the class of portable programs, we need to associate the identity of a process

with each occurrence of a shared variable in the process. (Section 4.1.1 explains this point in greater detail.) So we have tagged the meta-variables corresponding to processes ($proc$, p, pp), statements (ss) and expressions (e) with a subscript i representing the index of the associated process $proc_i$. Even though this aspect can be specified through the static semantics, we have chosen to make it explicit in the abstract syntax itself to emphasize (and facilitate) extensions required to deal with shared variables.

The set of processes has been partitioned into two groups: postponed processes (pp) and non-postponed processes (p). A process consists of a sequence of statements that can potentially be executed over and over again. The statements include (shared) variable assignments, signal assignments, wait statements, if statements and while statements. In wait statements, whenever on S, for e, or until e are omitted, on S_{ue} (where S_{ue} is the set of signals in the until clause), for ∞, or until $true$ respectively are assumed. In signal assignments, whenever the after-clause is omitted, after 0 is assumed. The expression syntax is standard (except for the one involving the $delayed$ signal attribute). Binary operators bop and unary operators uop are assumed to include standard logical operators (such as $\wedge, \vee, \neg$ etc) and arithmetic operators (such as $+, -, *$ etc).

3 Static Semantics

With regards to the static semantics, we assume that

- The programs are *well-typed*, that is, they satisfy all the usual type constraints.

- All the signals with multiple drivers have a suitable resolution function associated with them.

Furthermore, let the predicate *postponed?* be true of all postponed-process indices.

4 Structural Operational Semantics

We describe the semantics of the VHDL-93 subset by extending the work of Goossens [4]. Let *Val*, *Sig*, *Var*, *SVar*, *Expr*, and *SSt* denote the domains of values, signals, variables, shared variables, expressions and sequential statements respectively.

4.1 Semantic Entities

The state of a computation is captured by the history of values of each signal, the value bound to each variable and each shared variable, and the "activity" status of each postponed process.

Each process can be associated with a local store *LStore*. The domain *LStore* models the value bindings of the variables and the signals. Each variable holds either an integer value or a boolean value, that is, $Val = \mathcal{Z} \cup \mathcal{B}$. Each signal s is interpreted as a partial

function $f : \mathcal{Z} \mapsto Val_\perp$ satisfying the following constraints [4]: for $n < 0$, $f(n)$ is the value of the signal n time steps ago; $f(0)$ is the current value of the signal s; for $n \geq 0$, $f(n+1)$ is the projected value for n time steps into future. $f(1)$ contains the value scheduled for the next delta cycle. f contains at least $\langle -\infty, i \rangle$ and $\langle 0, v \rangle$ for initial value i and current value v of s. Note that only for $n > 0$ is $\langle n, \perp \rangle$ a valid pair in f and encodes a null transaction for time n.

The domain *SStore* models the value bindings of the shared variables. To guarantee portability of VHDL-93 programs, access to shared variables must be restricted. For instance, in any simulation cycle, all processes can read a shared variable, or exactly one process can read and write a shared variable, without jeopardizing portability. However, one cannot permit arbitrary reads and writes across processes. Thus, to characterize portable programs, we associate with each shared variable, its current value, the type of last access (read/write) and the index of the process accessing it. The distinguished constants $\perp$ and $\top$ represent *undefined* and *all* respectively. The precise details of the handling of shared variables will be given in Section 4.1.1.

It is also necessary to remember whether or not a postponed process is active, ready to be run at the end of the last delta cycle for the current time. Thus, the domain *PPStat* is modelled as a subset of (postponed) process indices I (or more precisely, a subset of *postponed?*).

The above discussion can be summarized by defining the signatures of the *semantic domains*. ($\mathcal{P}$ stands for the powerset operator.)

$$
\begin{aligned}
LStore &= (Var \mapsto Val) \times (Sig \mapsto \mathcal{P}(\mathcal{Z} \times Val_\perp)) \\
SStore &= (SVar \mapsto (Val \times (I \cup \{\perp, \top\}) \times \mathcal{P}(\{r, w\}))) \\
PPStat &= \mathcal{P}(I)
\end{aligned}
$$

4.1.1 Shared Variables

We motivate, through examples, the causes of non-portability. We assume that all integer variables/shared variables are initially 0.

Example 1.

$$\text{while true do } (sx := 1; \text{wait for } 1\ ns;)$$
$$\|$$
$$\text{while true do } (sx := 2; \text{wait for } 1\ ns;)$$

This program is not portable because the value of sx after t-ns (> 0) will be either 1 or 2 depending on how the simulator interleaves the computations of the two processes. Note that there are multiple concurrent writes to the shared variable.

Example 2.

$$\text{while true do } (y := y + 1; sx := y; \text{wait for } 1\ ns;)$$
$$\|$$
$$\text{while true do } (z := sx; \text{wait for } 1\ ns;)$$

This program is not portable because the value of z after t-ns will be either t or $t+1$ due to asynchronous execution. Note that one process writes to a shared variable and another process reads from it. This example also illustrates that the portability property is *not* local/compositional in nature (because a single process program is always portable).

Example 3. On the contrary, the following program is portable because, in each unit-time-interval, the shared variable is either only read simultaneously by both processes, or is accessed in read/write mode only by the second process.

$$\texttt{while true do } (y := sx; \texttt{wait for } \mathit{2 \ ns;})$$

$$\|$$

$$\texttt{while true do } (z := sx; \texttt{wait for } \mathit{1 \ ns}; \ sx := sx + 1; \texttt{wait for } \mathit{1 \ ns;})$$

Here the value of sx after t-ns is $\lceil \frac{t}{2} \rceil$.

Example 4. In general, the portability property cannot be checked statically (that is, at compile-time).

$$\texttt{while true do } (\texttt{if } \mathit{sflag} \texttt{ then } sx := \mathit{1} \texttt{ else } sx := \mathit{2}; \texttt{wait for } \mathit{2 \ ns;})$$

$$\|$$

$$\texttt{while true do } (sx := \mathit{1}; \texttt{wait for } \mathit{2 \ ns;})$$

The program has a unique meaning if and only if the boolean variable $sflag$ is always true. Note also that, when $sflag$ is true, the two processes write the same value into the shared variable sx. (We have assumed that the value of $sflag$ can be determined only at run-time.)

A deterministic finite-state automaton (DFA) is used to formally describe the "state" of the value bound to a shared variable. A DFA is a 5-tuple [6]: $(\mathbf{Q},\Omega,\Gamma,\mathbf{F}, q_0)$, where $\mathbf{Q}$ is the set of possible states, Ω is the alphabet, Γ is the transition function ($\Gamma : \mathbf{Q} \times \Omega \mapsto \mathbf{Q}$), $\mathbf{F}$ is the set of accepting states ($\subseteq \mathbf{Q}$), and q_0 is the initial state ($\in \mathbf{Q}$). We customize these sets for the problem at hand as follows:

- $\mathbf{Q} = Val \times (I \cup \{\perp, \top\}) \times \mathcal{P}(\{r, w\})$.

   Informally, the shared variable value is tagged with the type of access and the index of the process that accesses the value. The four possible types of accesses are: $\emptyset$, $\{r\}$, $\{w\}$ and $\{r, w\}$ representing *no access yet*, *read*-access, *write*-access, and *read/write*-access respectively. The $\perp$ value for the index signifies that no process has yet accessed the shared variable in the given unit-interval, while the $\top$ value means that all processes are allowed access.

   A note about the size of $\mathbf{Q}$. *The set I is finite, but the set Val is infinite, as defined. However, for our purposes, we make the simplifying and realistic assumption that Val is arbitrarily large but finite. (Overflow will trigger a run-time error.)*

- $\Omega = I \cup (I \times Val)$.

   The state of a shared variable changes when it is accessed. A read-action is represented by the index of the process from which the read has been issued, while a write-action

is represented by a pair consisting of the value to be written and the index of the process from which the write has been issued.

- $q_0 = \langle v, \perp, \emptyset \rangle$.

  v is the value of the shared variable at the beginning of the first simulation cycle for the current unit-interval. The index $\perp$ and the type of access $\emptyset$ signify that the shared variable has not yet been accessed.

- The deterministic transition function Γ is given below:

$$\langle v, \perp, \emptyset \rangle \xmapsto{i} \langle v, i, \{r\} \rangle \qquad\qquad \langle v, \perp, \emptyset \rangle \xmapsto{(i,u)} \langle u, i, \{w\} \rangle$$

$$\langle v, i, \{r\} \rangle \xmapsto{i} \langle v, i, \{r\} \rangle \qquad\qquad \langle v, i, \{r\} \rangle \xmapsto{j} \langle v, \mathsf{T}, \{r\} \rangle \quad \text{if } i \neq j$$

$$\langle v, i, \{r\} \rangle \xmapsto{(i,u)} \langle u, i, \{r, w\} \rangle \qquad \langle v, i, \{r\} \rangle \xmapsto{(j,u)} \langle u, \mathsf{T}, \{r, w\} \rangle \quad \text{if } i \neq j$$

$$\langle v, i, \{w\} \rangle \xmapsto{i} \langle v, i, \{r, w\} \rangle \qquad \langle v, i, \{w\} \rangle \xmapsto{j} \langle v, \mathsf{T}, \{r, w\} \rangle \quad \text{if } i \neq j$$

$$\langle v, i, \{w\} \rangle \xmapsto{(i,v)} \langle v, i, \{w\} \rangle \qquad \langle v, i, \{w\} \rangle \xmapsto{(i,u)} \langle u, i, \{r, w\} \rangle \quad \text{if } u \neq v$$

$$\langle v, i, \{w\} \rangle \xmapsto{(j,v)} \langle v, \mathsf{T}, \{w\} \rangle \quad \text{if } i \neq j \qquad \langle v, i, \{w\} \rangle \xmapsto{(j,u)} \langle u, \mathsf{T}, \{r, w\} \rangle \text{ if } i \neq j \wedge u \neq v$$

$$\langle v, i, \{r, w\} \rangle \xmapsto{i} \langle v, i, \{r, w\} \rangle \qquad \langle v, i, \{r, w\} \rangle \xmapsto{j} \langle v, \mathsf{T}, \{r, w\} \rangle \quad \text{if } i \neq j$$

$$\langle v, i, \{r, w\} \rangle \xmapsto{(i,u)} \langle u, i, \{r, w\} \rangle \qquad \langle v, i, \{r, w\} \rangle \xmapsto{(j,u)} \langle u, \mathsf{T}, \{r, w\} \rangle \quad \text{if } i \neq j$$

$$\langle v, \mathsf{T}, \{r\} \rangle \xmapsto{j} \langle v, \mathsf{T}, \{r\} \rangle \qquad\qquad \langle v, \mathsf{T}, \{r\} \rangle \xmapsto{(j,u)} \langle u, \mathsf{T}, \{r, w\} \rangle$$

$$\langle v, \mathsf{T}, \{w\} \rangle \xmapsto{j} \langle v, \mathsf{T}, \{r, w\} \rangle$$

$$\langle v, \mathsf{T}, \{w\} \rangle \xmapsto{(j,v)} \langle v, \mathsf{T}, \{w\} \rangle \qquad \langle v, \mathsf{T}, \{w\} \rangle \xmapsto{(j,u)} \langle u, \mathsf{T}, \{r, w\} \rangle \quad \text{if } u \neq v$$

$$\langle v, \mathsf{T}, \{r, w\} \rangle \xmapsto{j} \langle v, \mathsf{T}, \{r, w\} \rangle \qquad \langle v, \mathsf{T}, \{r, w\} \rangle \xmapsto{(j,u)} \langle u, \mathsf{T}, \{r, w\} \rangle$$

- $\mathbf{F} = (Val \times \{\perp\} \times \{\emptyset\}) \bigcup (Val \times I \times \{\{r\}, \{w\}, \{r, w\}\}) \bigcup (Val \times \{\mathsf{T}\} \times \{\{r\}, \{w\}\})$

  Informally, the set of accepting states characterizes the safe sequences of reads and writes in a portable program.

The states in $(Val \times \{\perp\} \times \{\{r\}, \{w\}, \{r, w\}\}) \cup (Val \times I \times \{\emptyset\}) \cup (Val \times \{\mathsf{T}\} \times \{\emptyset\})$ are *unreachable* from the start state, and the states in $Val \times \{\mathsf{T}\} \times \{\{r, w\}\}$ are the (dead or) absorbing states.

Lemma 4.1 *Every string (of actions) in the language $\mathcal{L}(\langle y, \mathsf{T}, \{r, w\} \rangle)$ contains a substring that matches one of the following regular expressions (where $i \neq j$ and $u \neq v$):*

  (a) $i\, j^*\, \langle j, v \rangle$ (b) $\langle i, v \rangle\, i^*\, (j \cup \langle j, u \rangle)$ (c) $\langle i, v \rangle\, \langle j, v \rangle^+\, (k \cup \langle k, u \rangle)$.

Proof: Follows by considering all the trasitions that cause the DFA to enter the state $\langle y, \mathsf{T}, \{r, w\} \rangle$ for the first time.   •

We now specify the (abbreviated) regular expressions corresponding to the sequence of reads and writes that take the DFA from the start state to each of the accepting states.

The "language" associated with each state is obtained by inspecting and reasoning about the transition function. (A word about the notation used here: "*" (resp. "+") represents Kleene star (resp. plus) operation. $?_{Set} \in Set$. Any two occurrences of $?_{Set}$ in a sequence may have different values.)

$$
\begin{aligned}
\mathcal{L}(\langle v, \perp, \emptyset \rangle) &= \{\epsilon\} \\
\mathcal{L}(\langle v, i, \{r\} \rangle) &= i^+ \\
\mathcal{L}(\langle u, i, \{w\} \rangle) &= \langle i, u \rangle^+ \\
\mathcal{L}(\langle u, i, \{r, w\} \rangle) &= \quad i \; \underline{(i \cup \langle i, ?_{Val} \rangle)^* \, \langle i, u \rangle \, i^*} \\
&\quad \bigcup \; \langle i, ?_{Val} \rangle^* \, \langle i, y \rangle \, \underline{(i \cup \langle i, ?_{Val} \rangle)^* \, \langle i, u \rangle \, i^*} \\
&\quad \bigcup \; \langle i, z \rangle^+ \, i \, \underline{(i \cup \langle i, ?_{Val} \rangle)^* \, \langle i, u \rangle \, i^*} \\
&\quad \bigcup \; \langle i, u \rangle^+ \, i^+ \qquad\qquad\qquad\qquad\qquad [u \neq y] \\
\mathcal{L}(\langle v, \top, \{r\} \rangle) &= i^+ \; j \; ?_I^* \qquad\qquad\qquad\qquad\qquad [i \neq j] \\
\mathcal{L}(\langle u, \top, \{w\} \rangle) &= \langle i, u \rangle^+ \, \langle j, u \rangle \, \langle ?_I, u \rangle^* \qquad\qquad [i \neq j]
\end{aligned}
$$

Lemma 4.2 *Every string (of read/write actions) in the language of the DFA satisfies one of the following properties:*

(a) *Every action in the string contains the same index i, that is, the action is either i or $\langle i, ?_{Val} \rangle$.*

(b) *Every action in the string is a read action, that is, it is in I.*

(c) *Every action in the string is a write action with the same value component, that is, it is in $I \times \{\{v\}\}$.*

Proof: By induction on the length of an accepting computation. •

Lemma 4.1 and Lemma 4.2 lay the foundation for defining portability.

Let $Size(rs)$ return the size of the set of indices in the read sequence rs. (For example, $size(ijkjijij) = 3$.)

Lemma 4.3 *Let $q, q_1, q_2 \in Q$, and $\longmapsto_*$ denote the reflexive transitive closure of $\longmapsto$. If $rs_1, rs_2 \in I^*$ are two sequences of reads such that $size(rs_1) > 1 \; \wedge \; size(rs_2) > 1$, then the relation $(q \overset{rs_1}{\longmapsto}_* q_1 \; \wedge \; q \overset{rs_2}{\longmapsto}_* q_2 \; \Rightarrow \; q_1 = q_2)$ holds.*

Proof: The result follows given that a read-transition has one of the following forms:

$$
\begin{aligned}
\langle v, \perp, \emptyset \rangle \overset{i}{\longmapsto} \langle v, i, \{r\} \rangle \qquad & \langle v, \top, A \rangle \overset{i}{\longmapsto} \langle v, \top, A \cup \{r\} \rangle \\
\langle v, i, A \rangle \overset{i}{\longmapsto} \langle v, i, A \cup \{r\} \rangle \qquad & \langle v, i, A \rangle \overset{j}{\longmapsto} \langle v, \top, A \cup \{r\} \rangle \quad \text{if } i \neq j \qquad \bullet
\end{aligned}
$$

Lemma 4.4 *Let $q, q_1, q_2 \in Q$, and $rs_1, rs_2 \in I^*$ be two sequences of reads that are permutations of each other. Then, the relation $(q \overset{rs_1}{\longmapsto}_* q_1 \wedge q \overset{rs_2}{\longmapsto}_* q_2 \; \Rightarrow \; q_1 = q_2)$ holds.*

Proof: We consider two cases: (a) $Size(rs_1) <= 1$. Trivial. (b) $Size(rs_1) > 1$. From Lemma 4.3. ●

4.1.2 Advancing time

A program is evaluated with respect to the global structure *Store* defined as follows:

$$Store = \mathcal{P}(LStore) \times SStore \times PPStat$$

$$\begin{array}{llll}
\sigma, \sigma_i & \in & LStore & \Sigma, \Sigma_i \in \mathcal{P}(LStore) \\
\psi & \in & SStore & \xi \in PPStat
\end{array}$$

Two functions — $\mathcal{T}, \mathcal{U} : Store \mapsto Store$ — are defined to advance time and delta time respectively, along the lines of [4]. (Note that *LStore* is referred to as *Store* in [4].)

The function $\mathcal{T}$ transforms a *Store* as follows:

- The (local) variables are unchanged: $\mathcal{T}(\sigma_i)(x) = \sigma_i(x)$.

- For signals: $\mathcal{T}(\sigma_i)(s) = \{\langle n-1, v\rangle \mid \langle n, v\rangle \in \sigma_i(s)\} \cup \{\langle 0, \sigma_i(s)(2) \text{ else } \sigma_i(s)(0)\rangle\}$.

  Here x *else* y means "if x is defined then x else y". Note that there is an error in [4] since it has 1 in place of 2, and as shown later, $\sigma_i(s)(1)$ is always undefined when $\mathcal{T}$ is applied.

- For shared variables: $\mathcal{T}(\psi)(sx) = \langle v, \perp, \emptyset\rangle$, where $\psi(sx) = \langle v, i, a\rangle$.

- For the status of the postponed-processes: $\mathcal{T}(\xi) = \emptyset$.

A signal s is *active* if $\exists \sigma_i \in \Sigma_I, v \in Val_\perp : \langle 1, v\rangle \in \sigma_i(s)$. A process can *resume* if it is sensitive to an active signal or it has been timed-out. (See Section 4.4.)

The function $\mathcal{U}$ effects only the value of the *active* signals and the status of the postponed processes. It leaves unchanged the values of variables, shared variables, and inactive signals.

- For active signals s, the current value is replaced by $r_s \in Val$, obtained through the signal resolution function f_s applied to the driving values of the signal [4]:

$$r_s = f_s\{\{v_i \mid \exists i \in I : \langle 1, v_i\rangle \in \sigma_i(s) \wedge v_i \neq \mathsf{null}\}\}$$
$$\mathcal{U}(\sigma_i)(s) = (\sigma_i(s) \setminus \{\langle 0, \sigma_i(s)(0)\rangle, \langle 1, \sigma_i(s)(1)\rangle\}) \cup \{\langle 0, r_s\rangle\}$$

  Here, $\{\{.\}\}$ denotes a multiset. f_s is assumed to be a commutative resolution function. null signifies disconnection. Note that inactive signals do not participate in determining the final resolved value.

- The status of the postponed processes is also updated by $\mathcal{U}$ as described later.

In the following three sections, we present the formal semantics of the VHDL-93 subset. The signatures of the relevant *semantic functions* are:

$$\mathcal{E} : Expr \mapsto LStore \times SStore \mapsto Val_\perp \times SStore$$
$$\rightarrow_{ss}, \rightarrow_{proc}: (LStore \times SStore \times SSt) \mapsto (LStore \times SStore \times SSt)$$
$$\rightarrow_{pgm} : (Store \times SSt) \times (Store \times SSt)$$

An expression is evaluated with respect to the local/shared store and it returns a value and a (possibly modified) shared store. A program (resp. statement) and a store evolve into a new program (resp. statement) and an (resp. unique) updated store.

4.2 Semantics of Expressions

Let *fst* stand for the function that extracts the first component of a pair and the set *dom(f)* stand for the domain of a partial function f. Let $\psi_v(sx) \in Val$ denote the first (value) component of the triple $\psi(sx)$ associated with the shared variable sx.

Also, $\psi[sx \mapsto st] = (\lambda sy.\ \text{if } sx \equiv sy \text{ then } st \text{ else } \psi(sy))$.

$$
\begin{array}{lll}
\mathcal{E}\,[\![\texttt{null}]\!]\,\langle\sigma,\psi\rangle & = & \langle null, \psi\rangle \\
\mathcal{E}\,[\![v]\!]\,\langle\sigma,\psi\rangle & = & \langle v, \psi\rangle \\
\mathcal{E}\,[\![x]\!]\,\langle\sigma,\psi\rangle & = & \langle \sigma(x), \psi\rangle \\
\mathcal{E}\,[\![sx_i]\!]\,\langle\sigma,\psi\rangle & = & \langle \psi_v(sx_i), \psi[sx_i \mapsto st]\rangle, \quad \text{if } \psi(sx_i) \overset{i}{\longmapsto} st \\
\mathcal{E}\,[\![s]\!]\,\langle\sigma,\psi\rangle & = & \langle \sigma(s)(0), \psi\rangle \\
\mathcal{E}\,[\![s'\texttt{delayed}(e_i)]\!]\,\langle\sigma,\psi\rangle & = & \langle \sigma(s)(n), \psi\rangle \quad n = max\{m \mid m \in dom(\sigma(s)) \land \\
& & \qquad\qquad m \le -\mathsf{fst}(\mathcal{E}\,[\![e_i]\!]\,\langle\sigma,\psi\rangle)\} \\
\mathcal{E}\,[\![uop\ e_i]\!]\,\langle\sigma,\psi\rangle & = & \langle uop\ v, \psi'\rangle \quad \text{if } \mathcal{E}\,[\![e_i]\!]\,\langle\sigma,\psi\rangle = \langle v, \psi'\rangle \\
\mathcal{E}\,[\![e_i\ bop\ e_i']\!]\,\langle\sigma,\psi\rangle & = & \langle v\ bop\ v', \psi''\rangle \quad \text{if } \mathcal{E}\,[\![e_i]\!]\,\langle\sigma,\psi\rangle = \langle v, \psi'\rangle \\
& & \qquad\qquad \text{and } \mathcal{E}\,[\![e_i']\!]\,\langle\sigma,\psi'\rangle = \langle v', \psi''\rangle
\end{array}
$$

$s'\texttt{delayed}(0\ ns) \ne s$ during any simulation cycle where there is a change in the value of s. (See Section 14.1 in the LRM [7].) For correct handling of delayed-attribute we also need to store the previous value of each signal in the *LStore*.

Theorem 4.1 *The meaning of an expression is independent of the order of evaluation of its subexpressions.*

Proof Sketch:

The meaning of an expression consists of its value and the shared store. As the expressions only inspect (read) the values bound to variables, shared variables and signals, and never modify (write) them, the value component is independent of the order of evaluation. By virtue of Lemma 4.4, permuting the sequence of reads on a shared variable has no effect on its final "state". So the result follows from structural induction. $\bullet$

As a consequence, the semantics does not change if we altered the last rule as follows:

$$
\begin{array}{lll}
\mathcal{E}\,[\![e_i\ bop\ e_i']\!]\,\langle\sigma,\psi\rangle & = & \langle v\ bop\ v', \psi''\rangle \quad \text{if } \mathcal{E}\,[\![e_i']\!]\,\langle\sigma,\psi\rangle = \langle v', \psi'\rangle \\
& & \qquad\qquad \text{and } \mathcal{E}\,[\![e_i]\!]\,\langle\sigma,\psi'\rangle = \langle v, \psi''\rangle
\end{array}
$$

4.3 Semantics of Statements

The semantic rules for all but the signal assignment statement and the wait statement are more or less standard, given that a process unwinds into a potentially infinite sequence of statements. (Recall that, $\sigma[x \mapsto v] = (\lambda y. \text{ if } x \equiv y \text{ then } v \text{ else } \sigma(y))$.)

$$\frac{true}{\langle \sigma, \psi, \ null \ ; \ ss \ \rangle \rightarrow_{ss} \langle \sigma, \psi, \ ss \ \rangle}$$

$$\frac{\mathcal{E} \ \llbracket e \rrbracket \ \langle \sigma, \psi \rangle = \langle v, \psi' \rangle}{\langle \sigma, \psi, \ x := e \ ; \ ss \ \rangle \rightarrow_{ss} \langle \sigma[x \mapsto v], \psi', \ ss \ \rangle}$$

$$\frac{\mathcal{E} \ \llbracket e \rrbracket \ \langle \sigma, \psi \rangle = \langle v, \psi' \rangle \quad \wedge \quad \psi'' = \psi'[\, sx_i \mapsto \Gamma(\psi'(sx_i), \langle i, v \rangle)\,]}{\langle \sigma, \psi, \ sx_i := e \ ; \ ss \ \rangle \rightarrow_{ss} \langle \sigma, \psi'', \ ss \ \rangle}$$

$$\frac{\langle \sigma, \psi, \ ss' \ \rangle \rightarrow_{ss} \langle \sigma', \psi', \ ss'' \ \rangle}{\langle \sigma, \psi, \ ss' \ ; \ ss \ \rangle \rightarrow_{ss} \langle \sigma', \psi', \ ss'' \ ; \ ss \ \rangle}$$

$$\frac{\mathcal{E} \ \llbracket e \rrbracket \ \langle \sigma, \psi \rangle = \langle \text{true}, \psi' \rangle}{\langle \sigma, \psi, \ \texttt{while} \ e \ \texttt{do} \ ss' \ \rangle \rightarrow_{ss} \langle \sigma, \psi', \ ss' \ ; \ \texttt{while} \ e \ \texttt{do} \ ss' \ \rangle}$$

$$\frac{\mathcal{E} \ \llbracket e \rrbracket \ \langle \sigma, \psi \rangle = \langle \text{false}, \psi' \rangle}{\langle \sigma, \psi, \ \texttt{while} \ e \ \texttt{do} \ ss' \ ; \ ss \rangle \rightarrow_{ss} \langle \sigma, \psi', \ ss \ \rangle}$$

$$\frac{\mathcal{E} \ \llbracket e \rrbracket \ \langle \sigma, \psi \rangle = \langle \text{true}, \psi' \rangle}{\langle \sigma, \psi, \ \texttt{if} \ e \ \texttt{then} \ ss' \ \texttt{else} \ ss'' \ \rangle \rightarrow_{ss} \langle \sigma, \psi', \ ss' \ \rangle}$$

$$\frac{\mathcal{E} \ \llbracket e \rrbracket \ \langle \sigma, \psi \rangle = \langle \text{false}, \psi' \rangle}{\langle \sigma, \psi, \ \texttt{if} \ e \ \texttt{then} \ ss' \ \texttt{else} \ ss'' \ \rangle \rightarrow_{ss} \langle \sigma, \psi', \ ss'' \ \rangle}$$

The signal assignment statement changes the value of a signal by adding an element (a time-value pair) and eliminating all other elements that are scheduled for a later time.

$$\frac{\mathcal{E} \ \llbracket e \rrbracket \ \langle \sigma, \psi \rangle = \langle v, \psi' \rangle \quad \wedge \quad \mathcal{E} \ \llbracket et \rrbracket \ \langle \sigma, \psi' \rangle = \langle t, \psi'' \rangle}{\langle \sigma, \psi, \ s \ \texttt{<=} \ e \ \texttt{after} \ et \ ; \ ss \rangle \rightarrow_{ss} \langle update(\sigma, s, v, t), \psi'', \ ss \ \rangle}$$

where $update(\sigma, s, v, t) = (\sigma(s) \setminus \{\langle n, \sigma(s)(n)\rangle \mid n > t\}) \cup \{\langle t+1, v\rangle\}$.

From Theorem 4.1, it also follows that the order of evaluation of the expressions e and et does not effect the final value of the signal s.

The rules for the wait-statements are given below in the definition of the semantic rules for the VHDL-93 programs.

4.4 Semantics of Processes and Programs

The semantic rules for processes/postponed processes (that is, for $\rightarrow_{proc}$) are similar to those for statements (that is, $\rightarrow_{ss}$).

A VHDL-93 program consists of a collection of sequential processes that execute independently. Global synchronization and (synchronous) communication through (common) signals takes place when all the processes reach a wait-statement. Otherwise, these processes execute asynchronously between wait-statements and can communicate (asynchronously) through shared variables, as captured by the following rule:(Here, we take the liberty of using the more intuitive $\|_I \langle \sigma_i, \psi, \xi, ss_i \rangle$ for $\langle \langle \|_I \sigma_i, \psi, \xi \rangle, \|_I ss_i \rangle$.)

$$\frac{\langle \sigma_j, \psi, \; ss_j \; \rangle \rightarrow_{ss} \langle \sigma_j', \psi', \; ss_j' \; \rangle}{\|_{I \cup \{j\}} \langle \sigma_i, \psi, \xi, \; ss_i \; \rangle \;\; \rightarrow_{pgm} \;\; \|_{I \cup \{j\}} \langle \sigma_i', \psi', \xi, \; ss_i' \; \rangle}$$

where $\sigma_i' = \sigma_i \;\wedge\; ss_i' = ss_i$ for all $i \neq j$, and $\sigma_i' = \sigma_j \;\wedge\; ss_i' = ss_j$ for $i = j$.

Observe that this rule causes nondeterministic execution of processes, and in general, the executions may yield different results in the presence of shared variables. However, for "portable" programs, this nondeterminism is tame and all executions are "equivalent". In the sequel, we use $ws_i[te_i, be_i]$ to stand for: (wait on S_i for te_i until be_i).

If no processes can resume, then the simulation time is advanced by one. To achieve this, the store is updated using $\mathcal{T}$ and the timeout value in the wait-statment is decremented by one. (Assume that $\infty - 1 = \infty$.)

$$\frac{\neg resume(\|_I \langle \sigma_i, \psi, \xi, \; ws_i[te_i, be_i] \;; ss_i \; \rangle) \;\; \wedge \;\; \forall \, i \in I : \; \langle tv_i, \psi' \rangle = \mathcal{E} \; [\![te_i]\!] \; \langle \sigma_i, \psi \rangle}{\|_I \langle \sigma_i, \psi, \xi, \; ws_i[te_i, be_i] \;; ss_i \; \rangle \;\; \rightarrow_{pgm} \;\; \|_I \langle \mathcal{T}(\sigma_i), \mathcal{T}(\psi), \mathcal{T}(\xi), \; ws_i[tv_i - 1, be_i] \;; ss_i \; \rangle}$$

This rule is well-defined because the order of evaluation of the te_i's and the function $\mathcal{T}$ does not modify the value (resp. the value component) of a variable (resp. a shared variable) or the sequence of values of a signal. (Note also that all the processes on the left hand side of $\rightarrow_{pgm}$ (in this rule and in all subsequent rules in this section) have a leading wait-statement.)

$$resume(\|_I \langle \sigma_i, \psi, \xi, \; ws_i[te_i, be_i] \;; ss_i \; \rangle) \;\; \equiv \;\; \exists i \in I : resume(\sigma_i, \psi, te_i) \;\; \vee \;\; (\xi \neq \emptyset)$$

A process can *resume* if it contains a signal that is active or it has been timed out.

$$resume(\sigma_i, \psi, te_i) \;\; \equiv \;\; active(\sigma_i) \;\vee\; timeout(\sigma_i, \psi, te_i)$$
$$active(\sigma) \;\; \equiv \;\; \exists s \in dom(\sigma), \; \exists v \in Val_\perp : \langle 1, v \rangle \in \sigma(s)$$
$$timeout(\sigma, \psi, te) \;\; \equiv \;\; fst(\mathcal{E} \; [\![te]\!] \; \langle \sigma, \psi \rangle) = 0$$

A delta cycle is initiated when a process can resume. Non-postponed processes are executed if they are timed-out or if the condition in the wait-statement holds, as shown below: (Recall also that VHDL-87 does not partition processes into postponed processes and non-postponed processes.)

$$\frac{\exists i \in I : \neg postponed?(i) \ \wedge \ resume(\sigma_i, \psi, te_i)}{\|_I \ \langle \sigma_i, \psi, \xi, \ ws_i[te_i, be_i] \,; ss_i \ \rangle \ \rightarrow_{pgm} \ \|_I \ \langle \mathcal{U}(\sigma_i), \psi, \xi', \ \mathcal{F}(ws_i[te_i, be_i] \,; ss_i) \ \rangle}$$

Informally, the function $\mathcal{F}$ executes the wait-statements for those non-postponed processes that are ready to run.

$$\mathcal{F}(ws_i[te_i, be_i] \,; ss_i) = \begin{cases} ss_i & \text{if } \neg postponed?(i) \ \wedge \ ready(\sigma_i, \mathcal{U}(\sigma_i), \psi, te_i, be_i) \\ ws_i[tv_i, be_i] \,; ss_i & \text{otherwise, where } tv_i = fst(\mathcal{E} \ [\![te_i]\!] \ \langle \sigma_i, \psi \rangle) \end{cases}$$

$$ready(\sigma_i, \sigma'_i, \psi, te_i, be_i) \ \equiv \ (\ timeout(\sigma_i, \psi, te_i) \ \vee$$
$$[\, \exists s \in S_i : event(\sigma_i, \sigma'_i, s) \ \wedge \ fst(\mathcal{E} \ [\![be_i]\!] \ \langle \sigma'_i, \psi \rangle) \,] \ )$$

Effectively, the timeout expression is evaluated only once in the first delta-cycle, while the condition in the wait-statement is evaluated in every delta cycle in which there is an event on a signal that the process/condition is "sensitive" to. (In other words, the timeout expression is evaluated in a store just prior to suspension, while the boolean condition is evaluated in a store just prior to reactivation [4].) Observe also that, in each delta cycle, it is determined whether or not a postponed process is ready to run (that is, it can be executed in the last delta cycle).

$$event(\sigma, \sigma', s) \ \equiv \ \sigma(s)(0) \neq \sigma'(s)(0)$$
$$\xi' \ \equiv \ \xi \ \cup \ \{i \in I \ \mid \ postponed?(i) \ \wedge \ ready(\sigma_i, \mathcal{U}(\sigma_i), \psi, te_i, be_i)\}$$

The postponed processes that can resume are executed only when no non-postponed process can resume. The condition that causes a postponed process to resume may no longer hold at the time the postponed process is actually executed. (See Section 8.1 in the LRM [7].) It is an error if the execution of a postponed process initiates another delta-cycle.

$$\frac{\begin{array}{c} \neg(\exists i \in I : \neg postponed?(i) \ \wedge \ resume(\sigma_i, \psi, te_i)) \ \wedge \ \xi \neq \emptyset \ \wedge \\ \forall i \in \xi : (\ \langle \mathcal{U}(\sigma_i), \psi, \ ss_i \ \rangle \rightarrow_{ss} \langle \sigma'_i, \psi', \ ws'_i[te'_i, be'_i] \,; ss'_i \ \rangle \ ) \ \wedge \\ \forall i \in I - \xi : (\ (\sigma'_i = \sigma_i) \ \wedge \ (ws_i[te_i, be_i] \,; ss_i \ \equiv \ ws'_i[te'_i, be'_i] \,; ss'_i) \ ) \\ \wedge \ \forall i \in I : \neg ready(\sigma'_i, \mathcal{U}(\sigma'_i), \psi', te'_i, be'_i) \end{array}}{\|_I \ \langle \sigma_i, \psi, \xi, \ ws_i[te_i, be_i] \,; ss_i \ \rangle \ \rightarrow_{pgm} \ \|_I \ \langle \sigma'_i, \psi', \emptyset, \ ws'_i[te'_i, be'_i] \,; ss'_i \ \rangle}$$

This rule is valid only because non-deterministic execution of the postponed processes of a "portable" VHDL-93 program result in unique values for shared variables.

We now define *portability* of VHDL-93 programs formally. Let $\rightarrow^*_{pgm}$ be the reflexive transitive closure of $\rightarrow_{pgm}$, and let $(\mathbf{Q}, \Omega, \Gamma, \mathbf{F}, q_0)$ be the DFA modelling the "states" of the shared variables.

Definition 4.1 *A program* $(\|_I$ while true do $ss_i)$ *is* portable *if, for every transition of the form* $(\|_I \ \langle \sigma_i, \psi, \xi, ss_i \rangle \ \rightarrow^*_{pgm} \ \|_I \ \langle \sigma'_i, \psi', \xi', ss'_i \rangle)$, *we have* $\psi' \in \mathbf{F}$.

5 Conclusions and Future Work

The designers of VHDL-93 found it useful to introduce shared variables into the language and stipulated that porgrams that are not portable are erroneous. We have investigated restriction/conditions under which programs with shared variables are portable.

In the immediate future, we plan to formally prove properties of portable VHDL-93 programs. We will also extend the language by introducing VHDL-93 constructs such as transport/inertial delays. Later we propose to design a simulator for the VHDL-93 subset that uses larger steps in simulation time.

References

[1] Bhasker, J., *A VHDL Primer*, Second Edition, Prentice Hall, Inc., 1994.

[2] Breuer, P., Sanchez, L., and Kloos, C. D., A simple denotational semantics, proof theory and validation condition generator for unit delay VHDL, *Formal Methods in System Design*, 7(1-2), July 1995.

[3] Kloos, C. D., and Breuer, P., eds., *Formal Semantics of VHDL*, vol. 307, Kluwer Academic Publishers, March 1995.

[4] Goossens, K. G. W., Reasoning about VHDL using operational and observational semantics, In: *Advanced Research Workshop on Correct Hardware Design Methodologies*, ESPRIT CHARME, Springer Verlag, October 1995.

[5] Hennessy, M., *The Semantics of Programming Languages: An Elementary Introduction using Structural Operational Semantics*, John Wiley & Sons, 1990.

[6] Hopcroft, J., and Ullman, J., *Introduction to Automata Theory, Languages and Computation*, Addison-Wesley Co, 1979.

[7] Institute of Electrical and Electronics Engineers, 345 East 47th Street, New York, USA. *IEEE Standard VHDL Language Reference Manual, Std 1076-1993*, 1993.

[8] van Tassel, J. P., *Femto-VHDL: The Semantics of a Subset of VHDL and its Embedding in the HOL Proof Assistant*, Ph. D. Dissertation, University of Cambridge, 1993.

[9] Wilsey, P. A., Developing a formal semantic definition of VHDL, In: Mermet, J., eds, *VHDL for Simulation, Synthesis and Formal Proofs of Hardware*, Kluwer Academic Publishers, pp. 243-256, 1992.

QUASI-STEADY STATE THERMAL RESISTANCE OF A FLEXIBLE COPPER-WATER HEAT PIPE SUBJECTED TO TRANSIENT ACCELERATION LOADINGS

Scott K. Thomas
Assistant Professor
Department of Mechanical and Materials Engineering

Wright State University
Dayton, OH 45435

QUASI-STEADY STATE THERMAL RESISTANCE OF A FLEXIBLE COPPER-WATER HEAT PIPE SUBJECTED TO TRANSIENT ACCELERATION LOADINGS

Scott K. Thomas
Assistant Professor
Department of Mechanical and Materials Engineering
Wright State University

Abstract

The thermal performance of a flexible copper-water heat pipe has been investigated to determine its quasi-steady state characteristics under varying acceleration loadings. This was accomplished by attaching the heat pipe to a centrifuge table, where the imposed angular velocity was sinusoidal in nature. It was found that the thermal resistance of the heat pipe is a function of the acceleration frequency, heat input, condenser temperature, and dryout condition prior to changing the frequency. The objective of the present experimental study is to determine the potential performance characteristics of heat pipes used as heat sinks in transient acceleration environments typical of those seen in high-performance aircraft. In addition, this research will enable heat pipe designers to re-examine the effects of acceleration loadings with respect to heat pipe wick and containment structures, so that new wicks and heat pipe shells can be developed and designed specifically for exploitation of the phenomena which occur in transient acceleration fields.

QUASI-STEADY STATE THERMAL RESISTANCE OF A FLEXIBLE COPPER-WATER HEAT PIPE SUBJECTED TO TRANSIENT RADIAL ACCELERATION LOADINGS

Scott K. Thomas

Introduction

The analysis of heat pipes has usually been restricted to the inclusion of a static acceleration field, such as that due to gravity. While this analysis is appropriate in many applications, it is not valid in the assessment of the thermal performance of heat pipes in acceleration fields which are varying with time. For instance, heat pipes have been proposed to be used aboard fighter aircraft such as the Navy F/A-18 to act as heat sinks for electronics packages which drive aileron or trailing edge flap actuators (Gernert et al., 1991; Yerkes and Beam, 1992; Yerkes and Hager, 1992). During combat, transient acceleration fields of up to 9-g's will be present on the aircraft. Therefore, knowledge of the thermal performance of heat pipes under transient acceleration loadings is of importance to designers of the electronics packages in need of cooling. The objective of the present experimental investigation is to determine the quasi-steady state thermal resistance of a flexible copper-water heat pipe under transient acceleration loadings with constant heat input. The performance of the heat pipe is examined for the following parameter ranges: Heat input, Q_e = 75 to 150 W; condenser temperature, T_c = 3, 20 and 35°C; and acceleration frequency, f = 0, 0.01, 0.05, 0.1, 0.15 and 0.2 Hz. The centrifuge radial acceleration loadings ranged from 1.1 to 9.8-g's for each frequency setting. In addition, the effects of the previous dryout history are noted.

Experimental Setup

Determination of the thermal resistance of the flexible copper-water heat pipe under transient radial acceleration loadings was accomplished by placing the pipe onto a 2.44 m diameter centrifuge table, as shown in Fig. 1. The heat pipe was mounted along a circular path at a radius of 1.149 m from the central axis of the centrifuge table. Transient acceleration fields were imposed on the heat pipe via open-loop control of the centrifuge

angular velocity by a waveform generator (Philips Model PM5139), which fed signals to the motor. The angular velocity of the centrifuge table was sinusoidal, $\omega = \sin ft$, but the radial acceleration was a sine-squared function given by $a_r = r\omega^2 = r\sin^2 ft$. The 20 hp dc electric motor was capable of maintaining a proper waveform on the table up to frequencies of $f = 0.2$ Hz, which dictated the upper end of the frequency testing. At higher frequencies, the response of the table was such that the imposed acceleration was attenuated from the desired range of 1.1 to 9.8-g's. The acceleration field at the heat pipe location was measured by a tri-axial accelerometer (Columbia Research Laboratories, Inc.) with an uncertainty of ± 0.1-g. Input power was supplied to the heat pipe from a precision power supply (Kepco Model ATE 150-3.5M) through power slip rings to the table. The input power was read by a power analyzer (Magtrol Model 4612B) which had an uncertainty of ± 1 W. A thermofoil heater (Minco Products) was attached to the evaporator mounting plate. Cooling fluid was delivered to the condenser mounting plate via a hydraulic rotary coupling (Dueblin, Inc.). The temperature of the cooling fluid was maintained at a constant setting by a recirculating chiller (Neslab Model HX-300) to within $\pm 0.5°$C. The mass flow rate of the coolant was measured by a calibrated rotameter (Dwyer Model RMA 24 SSV) to an uncertainty of ± 0.02 kg/min. Heat pipe temperatures were measured by Type T thermocouples ($\pm 0.5°$C). Temperature signals were amplified and conditioned on the centrifuge table. These signals were transferred off the table through instrumentation slip rings, which were completely separate from the power slip rings to reduce noise. Conditioning the temperature signals prior to leaving the centrifuge table eliminated difficulties associated with creating additional junctions within the slip ring assembly. Temperature and acceleration signals were collected using a personal computer and Keithley Viewdac data logging software. Information concerning the heat pipe is given in Table 1.

Test Procedure

The flexible copper-water heat pipe was tested in the following manner. The recirculating chiller was turned on and allowed to reach the setpoint temperature. The centrifuge table was started from the remote control room at a slow constant rotational speed ($f = 0$) to prevent damage to the power and instrumentation slip rings. Power to the heater was applied, and the heat pipe was allowed to reach a steady state condition, which was determined by monitoring various temperatures on the heat pipe. The centrifuge table angular velocity was then changed to a sinusoidal waveform. The frequency of the waveform driving the centrifuge table was increased in steps ($f = 0.01, 0.05, 0.1, 0.15, 0.2$ Hz), with steady conditions being reached at each setting. After all data had been recorded, the power to the heater was turned off, and the heat pipe allowed to cool before shutting down the centrifuge table. The temperature data was reduced and plotted using QuattroPro and Axum software. The thermal resistance of the heat pipe, defined as the evaporator-to-condenser temperature difference divided by the heat input, was calculated for each steady state condition

$$R_{th} = \frac{(T_e - T_c)}{Q_e} \qquad (1)$$

where T_e is the evaporator temperature. Calorimetric tests were performed, and it was determined that the difference between the heat input to the evaporator and that extracted by the condenser was less than 10%, so the heat input is used for data reduction. The maximum uncertainty of the reported thermal resistance data was calculated to be $\pm 9.5 \times 10^{-3}$ K/W. The repeatability of the data was examined, and the maximum difference in the thermal resistance observed between two identical runs was 7 %.

Results and Discussion

The performance characteristics of a flexible copper-water heat pipe under transient radial acceleration loadings have been determined using a centrifuge table. The thermal resistance of the heat pipe was measured while varying the following parameters: the acceleration frequency, evaporator heat input, and condenser temperature. The raw data

58-5

from a typical set of tests is shown in Fig. 2. The overall temperature difference across the heat pipe is slightly higher under the imposed transient acceleration field, as opposed to that for no acceleration ($f = 0$). Also, the temperature difference decreases as the centrifuge table frequency increases. This is a result of the fluid within the heat pipe sloshing from the condenser to the evaporator section, which effectively cools the evaporator by supplying additional fluid above that which is delivered by the capillary wick structure. For the lower heat inputs (Figs. 2(a) and 2(b)), the overall temperature difference does not change appreciably. For the higher heat input (Fig. 2(c)), the thermal resistance is significantly higher for the case of $f = 0.01$ Hz. At this point, the outboard evaporator pad temperature (closest to the evaporator end cap) exceeded the inboard temperature (closest to the adiabatic section), which is defined as a cross-over dryout condition. This is indicative of a partial dryout of the evaporator section wick structure, but due to the dynamics of the imposed acceleration field, it is not a catastrophic type of dryout normally encountered with stationary heat pipes. In other words, the heat pipe can continue to operate and reach a steady state condition under cross-over dryout conditions, even though the thermal resistance increases significantly. In some cases, the evaporator temperature reached $T_e = 100°C$, which is defined as an over-temperature dryout condition. Due to limitations of the instrumentation and due to the impracticality of allowing electronic components to reach such a high temperature, a steady-state over-temperature condition was not attempted. Table 2 presents the conditions under which the heat pipe experienced an over-temperature dryout condition. With no acceleration ($f = 0$), the heat pipe dried out at $Q_e = 150$ W for the lowest condenser temperature ($T_c = 3°C$). The heat input at which dryout occurred when the transient acceleration field was imposed decreased significantly for the lowest frequency settings. In these cases, the time period during which the heat pipe was tangentially accelerating and decelerating was relatively long. This resulted in a liquid pool which alternately resided in the condenser section and evaporator section for a significant amount of time. When the pool was in the condenser, the evaporator reached a partial dryout condition which was not recovered from when the liquid pool returned to the evaporator. The power input at which over-

temperature dryout occurred increased with the centrifuge table frequency for both condenser temperatures. Over-temperature dryout was not encountered for $T_c = 35°C$. Table 3 presents the resulting values of the thermal resistance of the heat pipe as a function of condenser temperature, heat input, and centrifuge table frequency. Values for the cases in which the frequency was both increased and decreased throughout the test are given. The test results for the stationary heat pipe ($f = 0$) are presented graphically in Fig. 3. The thermal resistance is nearly constant for condenser temperatures of $T_c = 20$ and $35°C$, but increases significantly with heat input after $Q_e = 113$ W for $T_c = 3°C$. This reduction in the capillary limit with operating temperature is well-documented in the heat pipe literature (Faghri, 1994; Dunn and Reay, 1982; Chi, 1976). The heat pipe thermal resistance for the case of increasing frequency is shown in Fig. 4. In general, the thermal resistance decreases as the table frequency increases, particularly above values of $f = 0.1$ Hz. This may be indicative of a natural frequency for fluid slosh within the heat pipe container. This particular phenomenon will be addressed in future analytical studies. The heat pipe thermal resistance also decreases as the condenser temperature increases. Therefore, the condenser section should reject heat to conditions which can maintain the operating temperature of the heat pipe significantly above the freezing temperature of the working fluid.

Partial Dryout Results

During the course of experimentation, it was discovered that the thermal resistance of the flexible copper-water heat pipe was dependent on the previous dryout conditions. This phenomenon is elucidated by examining the following set of tests, where the input power was $Q_e = 137$ W and the coolant temperature was $T_c = 20°C$. Figure 5(a) shows a typical test for the thermal characteristics of the flexible heat pipe (Test 1), where the centrifuge table frequency was increased throughout the test ($f = 0, 0.05, 0.1, 0.15, 0.2$ Hz). In Fig. 5(b), the table frequency was varied as follows: $f = 0, 0.15, 0.1, 0.05, 0.1$ Hz (Test 2). Figure 5(c) presents the results for Test 3, where the frequency variation was $f = 0, 0.2, 0.15, 0.1, 0.05, 0.1$ Hz. After the steady state was reached in Test 3 at $f = 0.1$ Hz, the

input power was shut off, and the heat pipe was allowed to cool and the evaporator wick to reprime fully before restarting the pipe. Table 4 summarizes the results of the three tests. The repeatability of the results are excellent, as can be seen in the evaporator temperatures for $f = 0.05$ Hz ($T_e = 97.0$, 99.4, 97.0°C). In addition, the transition from $f = 0.05$ to 0.1 Hz is also repeatable ($T_e = 75.1$, 74.2, 75.1°C for $f = 0.1$ Hz). Upon examination of the results for $f = 0.1$ Hz, it can be seen that the evaporator temperature is dependent on the previous temperature history of the heat pipe. For example, in Tests 1 and 3, the transition from $f = 0.05$ to 0.1 results in evaporator temperatures of $T_e = 75.1$°C at $f = 0.1$ Hz for each test. However, the transition from $f = 0.15$ to 0.1 Hz shown in Tests 2 and 3 gives $T_e = 54.7$ and 54.0°C, respectively (at $f = 0.1$ Hz). Similar results can be found for $f = 0.15$ Hz ($T_e = 67.8$°C for $f = 0.1$ to 0.15 Hz; $T_e = 42.3$°C for $f = 0.2$ to 0.15 Hz). In both cases, when the heat pipe did not experience a partial dryout at the previous frequency setting, the evaporator temperature was lower than when a partial dryout was present at the previous setting. In other words, if the evaporator temperature shows a partial dryout condition, increasing the frequency will decrease the evaporator temperature, but not as much as if no dryout was present. Therefore, the thermal resistance of the flexible copper-water heat pipe is dependent on the partial dryout status of the evaporator section prior to changing the frequency of the acceleration field.

Conclusions

The quasi-steady state thermal resistance of a flexible copper-water heat pipe under varying acceleration loadings has been determined experimentally. It was found that the thermal resistance of the heat pipe is a function of the sinusoidal frequency of the acceleration field, the heat input, the condenser temperature, and the dryout condition prior to changing the acceleration frequency. In order to determine the feasibility of using heat pipes in transient acceleration fields, the expected frequencies and heat inputs must be known. It was determined that increasing the condenser sink temperature will decrease the thermal resistance considerably by avoiding a partial dryout situation. While the

imposed acceleration field increased the heat pipe thermal resistance, the performance was improved at higher frequencies.

Recommendations for Future Work

Further testing is necessary to complete the data for the case of increasing frequency. Plans are being implemented to build and test axially-grooved and spirally-grooved heat pipes to examine the effects of different wick structures on the performance of the heat pipe under transient acceleration loadings. In addition, varying the nature of the waveform (sawtooth wave, square wave, etc.) is being considered.

References

1. Chi, S., 1976, *Heat Pipe Theory and Practice*, Hemisphere, Washington, D.C.

2. Dunn, P., and Reay, D., 1982, *Heat Pipes*, 3rd Edn., Pergamon, Oxford.

3. Gernert, N., et al., 1991, "Flexible Heat Pipe Cold Plates for Aircraft Thermal Control," *Proceedings of the Aerospace Technology Conference and Exposition*, SAE Paper No. 912105.

4. Faghri, A., 1994, *Heat Pipe Science and Technology*, Taylor and Francis, Washington, D.C.

5. Yerkes, K., and Beam, J., 1992, "Arterial Heat Pipe Performance in a Transient Heat Flux and Body Force Environment," *Proceedings of the Aerospace Atlantic Conference*, Dayton, OH, SAE Paper No. 921024.

6. Yerkes, K., and Hager, B., 1992, "Transient response of heat pipes for actuator thermal management," *Proceedings of the Aerospace Atlantic Conference*, Dayton, OH, SAE Paper No. 921024.

Table 1: Flexible heat pipe design specifications.

Heat pipe length	73.7 cm
Working fluid	Water
Working fluid charge	25 cm^3
Wall/wick materials	Copper
Evaporator envelope	OFHC waveguide tubing
Evaporator wick	+200/-325 sintered copper powder
Evaporator mounting plate	10.2 cm × 12.7 cm × 0.48 cm
Flexible artery	Braided cable
Bellows length	45.7 cm
Bellows inside diameter	1.3 cm
Condenser wick	Spiral grooves
Groove pitch	60 grooves per 2.54 cm
Groove depth	0.051 cm
Groove angle	30°
Condenser envelope	OFHC copper tubing
Tube outside diameter	1.59 cm
Tube wall thickness	0.102 cm
Condenser mounting plate	3.2 cm × 15.2 cm × 0.48 cm

Table 2: Over-temperature dryout conditions for the flexible copper-water heat pipe.

T_c (°C)	f (Hz)	Q_e (W)
3	0	≥150
3	0.01	≥113
3	0.05	≥125
3	0.1	≥137
3	0.15	≥150
3	0.2	≥150
20	0.01	≥125
20	0.05	≥150

Table 3: Thermal resistance of the flexible copper-water heat pipe (OTD = Over-Temperature Dryout).

T_c (°C)	Q_e (W)	f (Hz)	R_{th} (K/W) Increasing f	R_{th} (K/W) Decreasing f
3	75	0	8.96×10^{-2}	-
3	75	0.01	2.37×10^{-1}	-
3	75	0.05	2.34×10^{-1}	-
3	75	0.1	1.19×10^{-1}	-
3	75	0.15	1.16×10^{-1}	-
3	75	0.2	1.11×10^{-1}	-
3	87	0	7.63×10^{-2}	-
3	87	0.01	4.31×10^{-1}	-
3	87	0.05	3.83×10^{-1}	-
3	87	0.1	1.96×10^{-1}	-
3	87	0.15	1.55×10^{-1}	-
3	87	0.2	1.25×10^{-1}	-
3	100	0	8.33×10^{-2}	-
3	100	0.01	6.17×10^{-1}	-
3	100	0.05	5.36×10^{-1}	-
3	100	0.1	3.62×10^{-1}	-
3	100	0.15	3.10×10^{-1}	-
3	100	0.2	2.88×10^{-1}	-
3	113	0	-	7.24×10^{-2}
3	113	0.01	-	OTD
3	113	0.05	-	6.81×10^{-1}
3	113	0.1	-	2.83×10^{-1}
3	113	0.15	-	2.22×10^{-1}
3	113	0.2	-	9.04×10^{-2}
3	125	0	-	2.27×10^{-1}
3	125	0.01	-	OTD
3	125	0.05	-	OTD
3	125	0.1	-	5.55×10^{-1}
3	125	0.15	-	4.83×10^{-1}
3	125	0.2	-	3.87×10^{-1}
3	137	0	-	3.60×10^{-1}
3	137	0.01	-	OTD
3	137	0.05	-	OTD
3	137	0.1	-	OTD
3	137	0.15	-	5.04×10^{-1}
3	137	0.2	-	4.86×10^{-1}

Table 3, cont.: Thermal resistance of the flexible copper-water heat pipe (OTD = Over-Temperature Dryout).

T_c (°C)	Q_e (W)	f (Hz)	R_{th} (K/W) Increasing f	R_{th} (K/W) Decreasing f
20	75	0	4.89×10^{-2}	-
20	75	0.01	8.5×10^{-2}	-
20	75	0.05	8.53×10^{-2}	-
20	75	0.1	7.21×10^{-2}	-
20	75	0.15	6.91×10^{-2}	-
20	75	0.2	6.71×10^{-2}	-
20	87	0	4.33×10^{-2}	-
20	87	0.01	7.84×10^{-2}	-
20	87	0.05	8.00×10^{-2}	-
20	87	0.1	6.89×10^{-2}	-
20	87	0.15	6.62×10^{-2}	-
20	87	0.2	6.51×10^{-2}	-
20	100	0	4.05×10^{-2}	-
20	100	0.01	2.1×10^{-1}	-
20	100	0.05	1.88×10^{-1}	-
20	100	0.1	6.81×10^{-2}	-
20	100	0.15	6.69×10^{-2}	-
20	100	0.2	6.3×10^{-2}	-
20	113	0	4.21×10^{-2}	-
20	113	0.01	3.61×10^{-1}	-
20	113	0.05	2.91×10^{-1}	-
20	113	0.1	1.27×10^{-1}	-
20	113	0.15	9.88×10^{-2}	-
20	113	0.2	8.40×10^{-2}	-
20	125	0	4.21×10^{-2}	-
20	125	0.01	OTD	-
20	125	0.05	3.75×10^{-1}	-
20	125	0.1	2.31×10^{-1}	-
20	125	0.15	-	-
20	125	0.2	-	-
20	137	0	4.23×10^{-2}	4.01×10^{-2}
20	137	0.01	OTD	OTD
20	137	0.05	4.67×10^{-1}	4.85×10^{-1}
20	137	0.1	2.96×10^{-1}	1.42×10^{-1}
20	137	0.15	2.39×10^{-1}	5.69×10^{-2}
20	137	0.2	2.31×10^{-1}	5.69×10^{-2}

Table 3, cont.: Thermal resistance of the flexible copper-water heat pipe (OTD = Over-Temperature Dryout).

T_c (°C)	Q_e (W)	f (Hz)	R_{th} (K/W) Increasing f	R_{th} (K/W) Decreasing f
20	150	0	-	3.74×10^{-2}
20	150	0.01	-	OTD
20	150	0.05	-	OTD
20	150	0.1	-	1.93×10^{-1}
20	150	0.15	-	1.72×10^{-1}
20	150	0.2	-	1.66×10^{-1}
35	75	0	4.00×10^{-2}	-
35	75	0.01	5.95×10^{-2}	5.92×10^{-2}
35	75	0.05	5.44×10^{-2}	5.41×10^{-2}
35	75	0.1	4.53×10^{-2}	4.57×10^{-2}
35	75	0.15	4.37×10^{-2}	4.71×10^{-2}
35	75	0.2	4.35×10^{-2}	4.53×10^{-2}
35	87	0	3.97×10^{-2}	3.79×10^{-2}
35	87	0.01	5.86×10^{-2}	5.78×10^{-2}
35	87	0.05	5.40×10^{-2}	5.33×10^{-2}
35	87	0.1	4.52×10^{-2}	4.52×10^{-2}
35	87	0.15	4.29×10^{-2}	4.26×10^{-2}
35	87	0.2	4.21×10^{-2}	4.21×10^{-2}
35	100	0	3.93×10^{-2}	-
35	100	0.01	5.81×10^{-2}	-
35	100	0.05	5.36×10^{-2}	-
35	100	0.1	4.49×10^{-2}	-
35	100	0.15	4.24×10^{-2}	-
35	100	0.2	4.17×10^{-2}	-
35	113	0	3.54×10^{-2}	-
35	113	0.01	5.67×10^{-2}	-
35	113	0.05	5.27×10^{-2}	-
35	113	0.1	4.23×10^{-2}	-
35	113	0.15	4.09×10^{-2}	-
35	113	0.2	4.05×10^{-2}	-
35	125	0	2.98×10^{-2}	-
35	125	0.01	1.11×10^{-1}	-
35	125	0.05	8.11×10^{-2}	-
35	125	0.1	4.23×10^{-2}	-
35	125	0.15	4.14×10^{-2}	-
35	125	0.2	4.12×10^{-2}	-

Table 3, cont.: Thermal resistance of the flexible copper-water heat pipe (OTD = Over-Temperature Dryout).

T_c (°C)	Q_e (W)	f (Hz)	R_{th} (K/W) Increasing f	R_{th} (K/W) Decreasing f
35	137	0	3.17×10^{-2}	-
35	137	0.01	2.23×10^{-1}	-
35	137	0.05	1.34×10^{-1}	-
35	137	0.1	5.00×10^{-2}	-
35	137	0.15	4.15×10^{-2}	-
35	137	0.2	4.12×10^{-2}	-
35	150	0	-	3.03×10^{-2}
35	150	0.01	-	3.23×10^{-1}
35	150	0.05	-	1.32×10^{-1}
35	150	0.1	-	4.15×10^{-2}
35	150	0.15	-	3.99×10^{-2}
35	150	0.2	-	3.99×10^{-2}

Table 4: Summary of partial dryout test results (Q_e = 137 W, T_c = 20°C).

Test 1		Test 2		Test 3	
f (Hz)	T_e (°C)	f (Hz)	T_e (°C)	f (Hz)	T_e (°C)
0	42.8	0	42.5	0	42.8
0.05	97.0	0.15	42.3	0.2	42.3
0.1	75.1	0.1	54.7	0.15	42.3
0.15	67.8	0.05	99.4	0.1	54.0
0.2	66.6	0.1	74.2	0.05	97.0
-	-	-	-	0.1	75.1
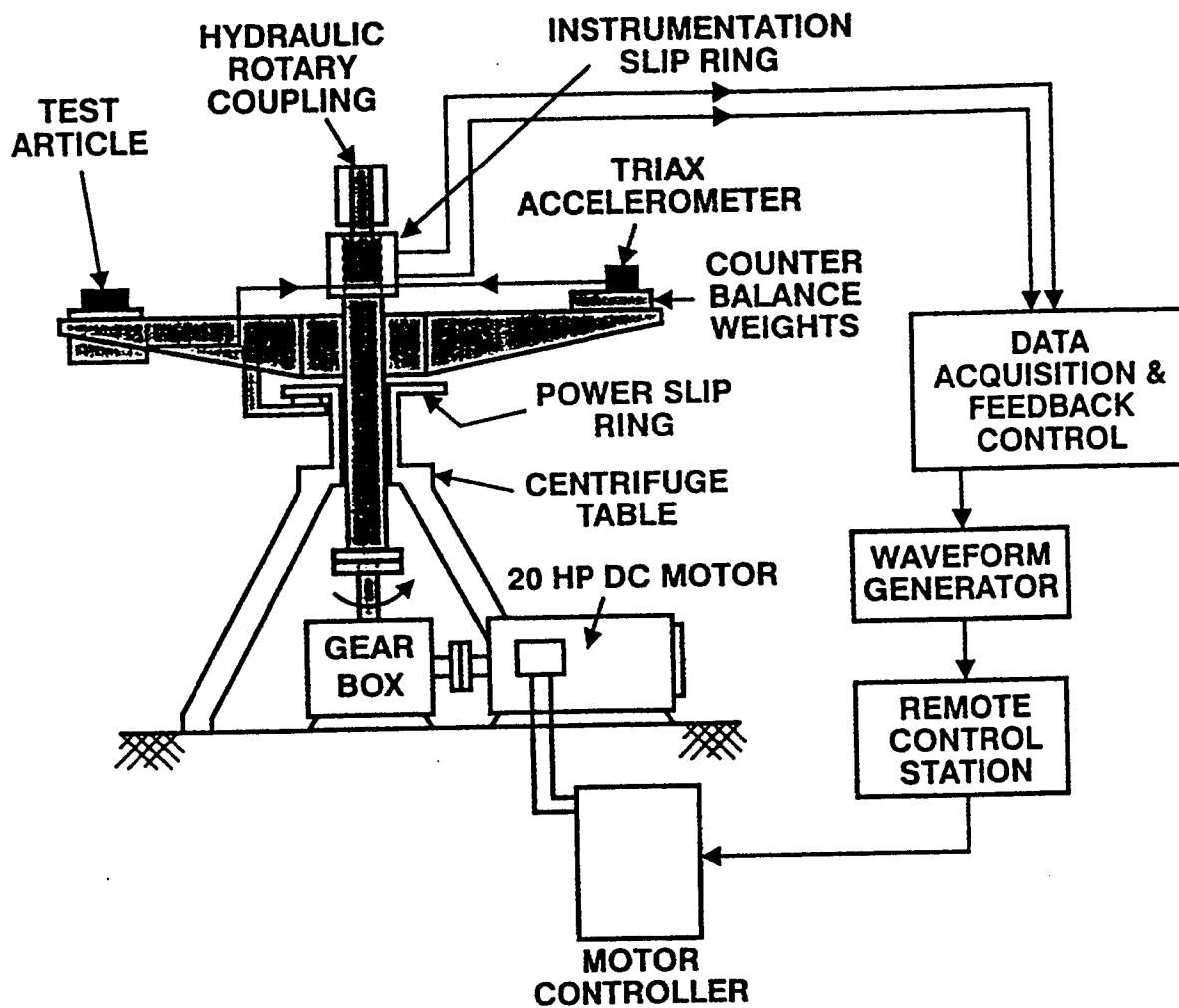
Figure 1: Schematic of the test apparatus.
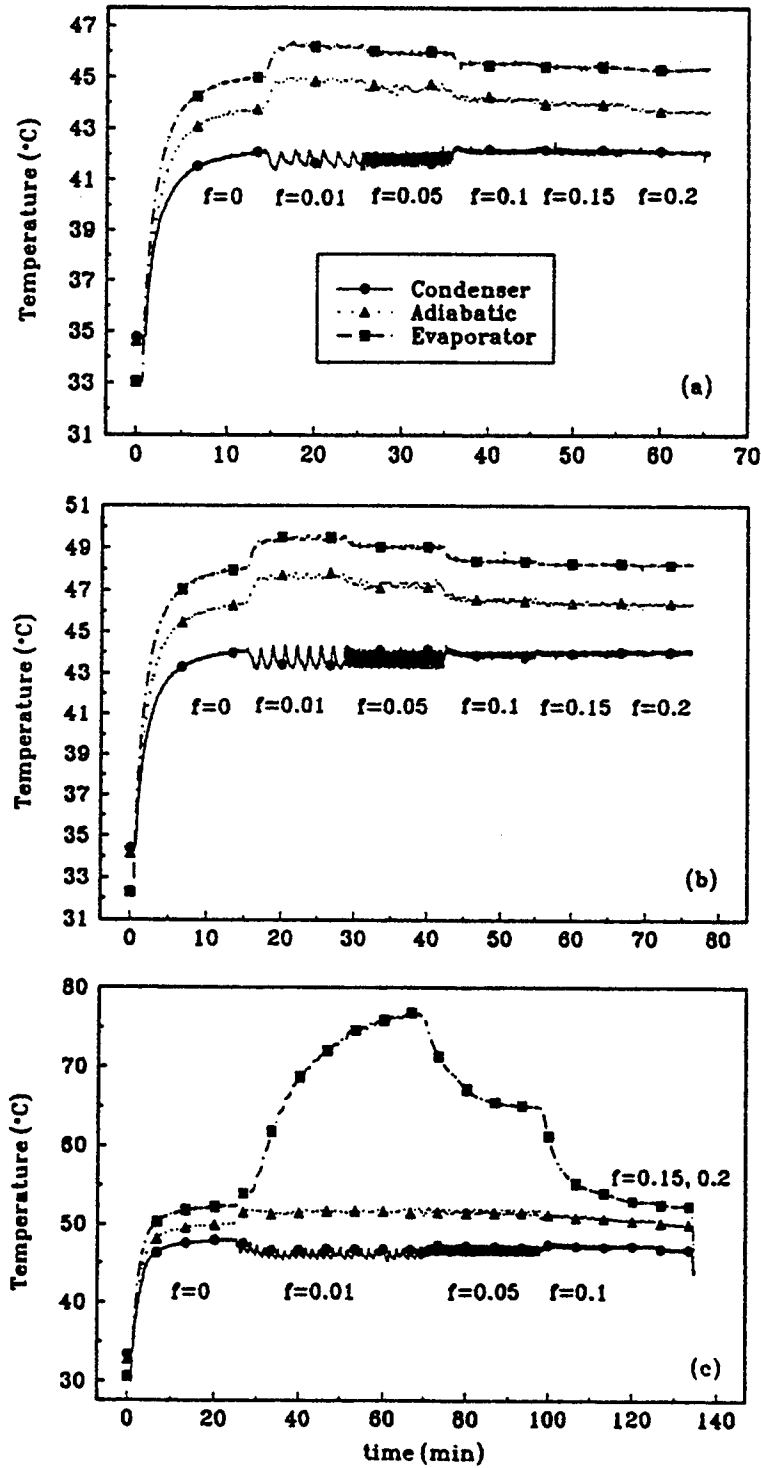
Figure 2: Temperature versus time for $T_c = 35°C$ and various frequency settings: (a) $Q_e = 75$ W; (b) $Q_e = 100$ W; (c) $Q_e = 137$ W.
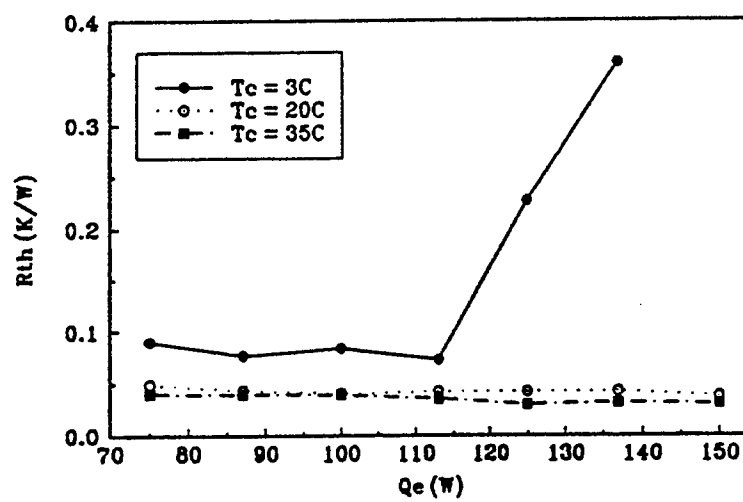
Figure 3: Thermal resistance versus heat input for the stationary heat pipe ($f = 0$).
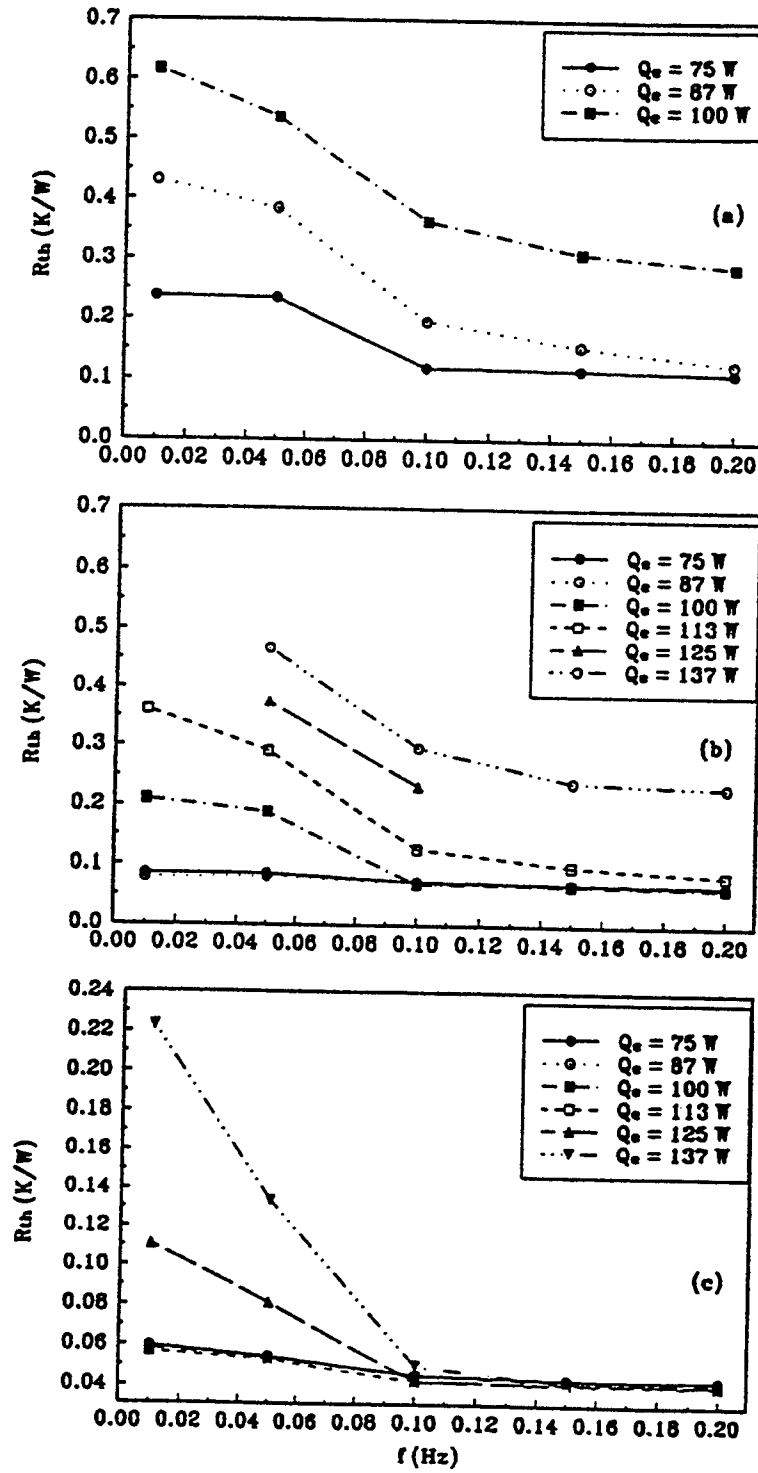
Figure 4: Thermal resistance versus heat input for increasing f: (a) $T_c = 3°C$; (b) $T_c = 20°C$; (c) $T_c = 35°C$.
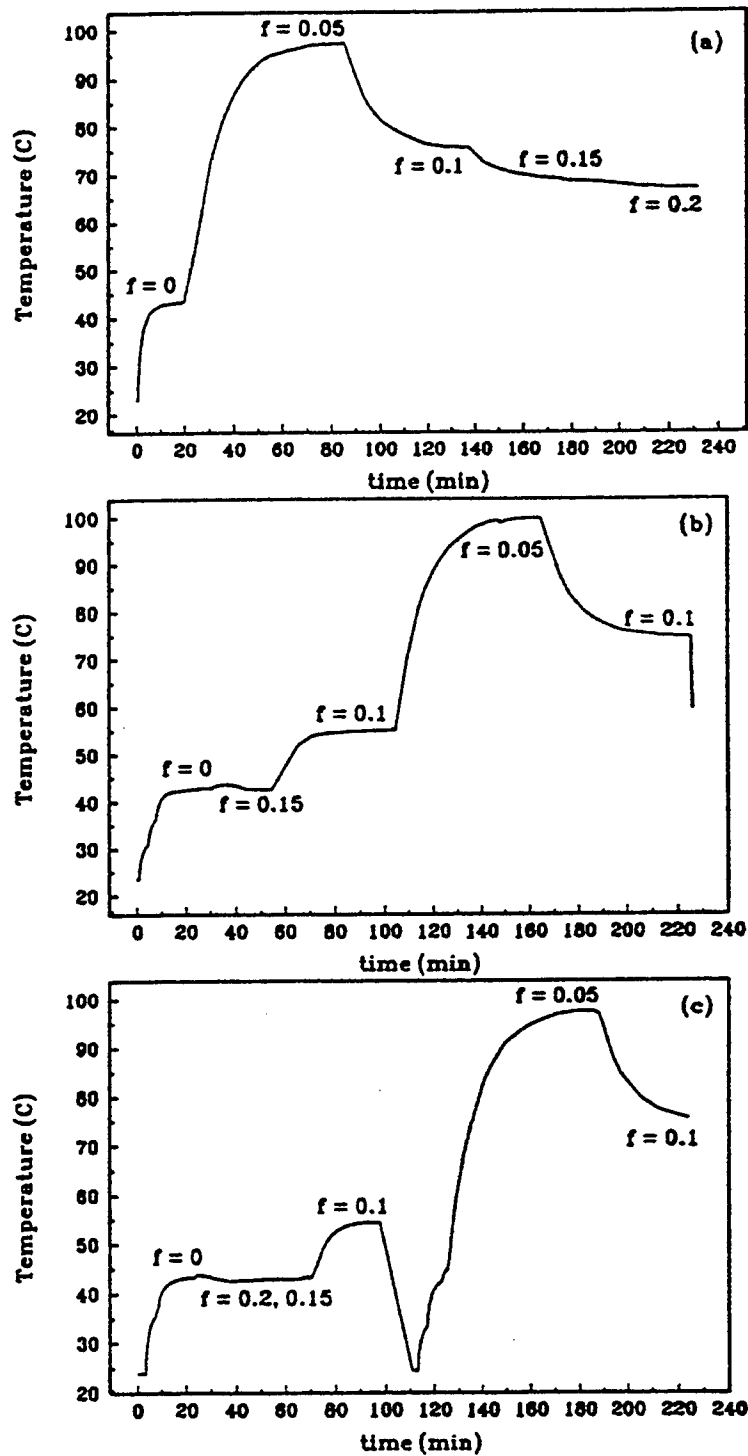
Figure 5: Temperature versus time for the heat pipe under different dryout conditions: (*a*) Test 1, increasing frequency; (*b*) Test 2, decreasing, then increasing; (*c*) Test 3, decreasing, allowing to reprime, then increasing.

A STUDY OF THE IMPLEMENTATION OF RESEARCH EPIC ON CRAY-T3D MASSIVELY PARALLEL COMPUTER USING PVM

C. T. Tsai
Associate Professor
Department of Mechanical Engineering

Florida Atlantic University
777 Glades Road
Boca Raton, FL 33431

Final Report for:
Summer Faculty Research Program
Wright Laboratory

August 1995

A STUDY OF THE IMPLEMENTATION OF RESEARCH EPIC ON CRAY-T3D MASSIVELY PARALLEL COMPUTER USING PVM

C. T. Tsai
Associate Professor
Department of Mechanical Engineering
Florida Atlantic University

Abstract

Research EPIC code can be mostly parallelized except the slide interface algorithm. A PVM EPIC code which executes the parallel algorithm of EPIC in many PEs of T3D while executes the sequential algorithm in one PE of T3D or YMP will improve the speedup of the whole EPIC code. The PVM EPIC will also be a portable code which can be implemented to any cluster of computer network. A initial study is performed by executing 2 paralleled subroutines (SOLID and VOLUME) on the T3D while executing the rest of the EPIC code on either the YMP or one PE of T3D. The results obtained from this initial study show that for an example having 40% of the code being paralleled, the CPU time for running the new PVM EPIC in T3D is about 2 times more than the one for running the sequentrial EPIC in YMP. It indicates that this PVM EPIC in T3D can run as fast as the sequentrial EPIC in YMP if 80% of the code are executed in parallel. When 90% of the code are executed in parallel, the PVM EPIC in T3D can double the speed of sequential EPIC in YMP. This goal can be accomplished by converting more EPIC subroutines into parallel program. This study encourage continuous works to parallel more EPIC subroutines and implement into the PVM EPIC to enhance the high performance computing of EPIC.

A STUDY OF THE IMPLEMENTATION OF RESEARCH EPIC ON CRAY-T3D MASSIVELY PARALLEL COMPUTER USING PVM

C. T. Tsai

Introduction

PVM (parallel Virtual Machine) is a software system to enable a collection of heterogeneous computer systems to be used as a coherent and flexible concurrent computational resource. It is a subroutine library that supports parallel programming through message-passing which is an explicit method to specifically request data be sent from one task to another, or between groups of tasks.

For CRAY T3D systems coupled with CRAY Y-MP system, it creates an idea system for PVM. This coupled system can be used to program on the CRAY T3D system alone or to let CRAY T3D communicate with processes running on the CRAY Y-MP system or any of the other systems that run PVM. For the PVM run on the CRAY T3D alone, the sequential portion of EPIC code was run on one PE while the parallel portion of the EPIC was run on all the PEs in T3D. For the PVM run on the T3D and YMP, the sequential portion of EPIC code was run on the YMP while the parallel portion of the EPIC was run on the T3D. Runing PVM on T3D alone spends less time on message-passing than that on the coupled T3D and YMP, but runing sequential portion of EPIC on one PE is slower than that on the YMP. Which method giving better performance depends on the amount of data being passed between sequential and parallel portion of EPIC code. Even the performance of applying both methods to the whole EPIC code

can not be predicted until they are completely finished and tested, we will apply both methods to 2 subroutines of EPIC code to test the validity of both methods.

In the first method, the EPIC code was run on the Cray YMP except the two subroutines VOLUME and SOLID which were converted to massively parallel program and executed on the Cray T3D. In the second method, the EPIC code was run on one PE of T3D except the two subroutines VOLUME and SOLID which were converted to massively parallel program and executed on the Cray T3. It was accomplished by using the message passing primitives of PVM. Of the various versions of PVM, the network version of PVM was used in this case. The approach used in this case was structured so that other subroutines can be added in the list of T3D subroutines with minimum changes. The methodology for both methods are the same and discussed next.

Methodology

The structure of the methodology is discussed below:

Makefile

The Makefile is created primarily to compile the programs on respective machines and create executables. The programs epic94t.f, solid_sp3.f,volume_sp3.f and xfercomt.f were compiled in T3D and the rest on YMP or PE0 on T3D.The executables that were created are epic94 and epic94t. The -dp was used in the compiling flag to disable the double precision arithmatic on YMP. The Makefile is run by batch process which is in the file makeepic.

DATA TRANSFER

Xfercom.f

The xfercom.f is the program which handles the communication or data transfers between the T3D subroutines and the YMP subroutines. The xfercom.f is build of various subroutines which are called by the T3D programs - in the present case by volume_sp3.f and solid_sp3.f. The subroutine transfers the data which are shared and are in include files or common block.

For example, the variables in MATERL are handled by subroutine XMATERL and XMATERLD. In the subroutine XMATERLD, first the data is shared. It may be called in four different ways - from child process, from parent process, to child process and to parent. On being called by any of the alternative way, it performs a sequence of operations. If called by the parent it initializes the buffer packs the data in the buffer and sends to the child process.

```
      SUBROUTINE XMATERLD (TOFROM, METHOD)

      INTEGER TOFROM
      INTEGER METHOD

      INCLUDE 'fpvm3.h'
      INCLUDE 'EPICPV'
      INCLUDE 'MATERL'

CDIR$ GEOMETRY GZ(:BLOCK(1))
CDIR$ SHARED (GZ) :: AM,AN,DEN,G,PMIN,SPH,TEMP1,TMELT
CDIR$ SHARED (GZ) :: TROOM,TZERO,IDAM,IFAIL,MODEL,MTYPE
```

```
      REAL PVMRBUF(MXMAT)
        INTEGER PVMIBUF(MXMAT)
      IF (TOFROM .EQ. TOCHILD) THEN
C SEND MATERL
        CALL PVMFINITSEND (PVMDATADEFAULT,ISBUF)
        IF (ISBUF .le. 0) then
          WRITE (6,*) 'Error on initsend in MATERL - exit'
          call exit (1)
        endif

        CALL PVMFPACK (REAL8, AM, MXMAT, 1, IPVMST)
        CALL PVMFPACK (REAL8, AN, MXMAT, 1, IPVMST)
        CALL PVMFPACK (REAL8, DEN, MXMAT, 1, IPVMST)
        ...

        CALL PVMFPACK (BYTE1, MTYPE, MXMAT*INTSIZ, 1, IPVMST)
        CALL PVMFSEND (PVMTIDS(1), SMATERL, IPVMST)
        IF (IPVMST .NE. PVMOK) THEN
          WRITE(6,*) 'ERROR ON SENDING MATERL - EXIT'
          CALL EXIT (1)
        ENDIF
      ELSE IF (TOFROM .EQ. FROMPARENT) THEN
        ...
      RETURN
      END
```

epic94.f

The epic94.f is the main program which starts the code on YMP. In terms of PVM, it initializes various PVM 'parameters'. It opens files for storing the data transfers for the T3D subroutines like for VOLUME , it opens volume.xfr file. Then, it spawns the process to the T3D PEs. Then,

it sends using the pvmfmcast (multicast) the variables XFRTYPE, VPATHNAME, VPATHLEN

etc. (PVM parameters defined in the epic94.f program).

```
PROGRAM EPIC94

    ...

EPICKYW = 'EPIC_METHOD'

EPICKYWL = 0

EPICVAL = ' '

EPICVALL = 0

CALL PXFGETENV (EPICKYW,EPICKYWL,EPICVAL,EPICVALL,IER)

IF (IER .NE. 0 .OR. EPICVALL .EQ. 0) THEN

    XFRTYPE = USEDISK

ELSE IF (EPICVAL(1:EPICVALL) .EQ. 'USEPVM') THEN

    XFRTYPE = USEPVM

    ...

    IF (IER .NE. 0 .OR. EPICVALL .EQ. 0) THEN

        VPATHNAME = 'VOLUME.xfr'

        VPATHLEN = LEN ('VOLUME.xfr')

    ELSE

        VPATHNAME = EPICVAL(1:EPICVALL)//'/'//'VOLUME.xfr'

        VPATHLEN = EPICVALL + 1 + LEN ('VOLUME.xfr')

        WRITE (6,*) VPATHNAME

    ENDIF

    VOLUMFN = VOLUMFNP

CALL ASSIGN ('assign -s unblocked f:'//VPATHNAME(1:VPATHLEN))


OPEN(VOLUMFN,FILE=VPATHNAME(1:VPATHLEN),STATUS='SCRATCH',

1       FORM='UNFORMATTED',ACCESS='SEQUENTIAL',

2       IOSTAT=IOS)
```

```
        IF (IOS .NE. 0) THEN
          WRITE (6,*) 'ERROR OPENING VOLUME XFR FILE - EXIT'
          CALL EXIT (1)
        ENDIF

        ...

        CALL PVMFSPAWN ('epic94t',
      &            PVMARCH,
      &            'CRAY',
      &            NUMTASKS,
      &            PVMTIDS,
      &            NTASK)

        ...

        CALL PVMFINITSEND (PVMDATADEFAULT,ISBUF)

        ...

        CALL PVMFPACK (BYTE1, XFRTYPE, INTSIZ, 1, IPVMST)
        CALL PVMFPACK (BYTE1, VPATHLEN, INTSIZ, 1, IPVMST)
        CALL PVMFPACK (BYTE1, SPATHLEN, INTSIZ, 1, IPVMST)

        ...

        END
```

<u>epic94t.f</u>

It first checks for the parent tid. It is the file which is spawned by the epic94.f. It receives
(unblocking receive) the data for XFRTYPE, VPATHLEN, SPATHLEN, VPATHNAME and
SPATHNAME. It opens files VOLUMFN and SOLIDFN which are the xfr files. The data for
material properties are also present in the file.

```
        PROGRAM EPIC94T

          ...
```

```fortran
      call pvmfparent(prntid)
      ...
         CALL PVMFRECV (PRNTID, -1, IRBUF)
         CALL PVMFBUFINFO (IRBUF, IBUFSZ, IRMSG, IRTID, IPVMST)
         IF (IPVMST .LT. 0) THEN
            WRITE (6,*) 'BAD PVMBUFINFO ON PARENT RECEIVE - EXIT'
            CALL EXIT (1)
         ENDIF
         IF (IRMSG .EQ. SHTDWN) THEN
            write (6,*) 'Shutdown message received'
            call flush (6)
            SHUTDOWN = .TRUE.
         ELSE IF (IRMSG .EQ. INITCHILD) THEN
            CALL PVMFUNPACK (BYTE1, XFRTYPE, INTSIZ, 1, IPVMST)
            CALL PVMFUNPACK (BYTE1, VPATHLEN, INTSIZ, 1, IPVMST)
            CALL PVMFUNPACK (BYTE1, SPATHLEN, INTSIZ, 1, IPVMST)
            VPATHNAME = ' '
      ...
         ELSE IF (IRMSG .EQ. CVOLUM) THEN
C            write (6,*) 'CVOLUM message received'
C            call flush (6)
            CALL VOLUME
```

volume_sp3.f

It is one of the subroutines which are compiled on the T3D. Firstly, the data is shared using the CRAFT shared directives. It is called by epic94t.f. On being called, it receives the data to be used by it from the parent process. Then it unpacks the data or is read from the xfr file called VOLUMFN. Then, the non array data like L1, LN etc. are copied to other PEs. It then calls the

subroutines XMISC which is in the xfercom.f (this is for the case FROMPARENT and USEPVM). It now has the shared data to run the program on T3D and so then calls the subroutine VOLUME_T3D, which is the shared to private conversion code for the original VOLUME subroutine. On finishing the VOLUME_T3D subroutine, again the data is transferred by packing and sending it to the parent.

```
      SUBROUTINE VOLUME
         ...
CDIR$ GEOMETRY GZ(:BLOCK(1))
CDIR$ SHARED (GZ) :: X1,X2,X3,X4

         ...
CDIR$ MASTER
      CALL PVMFRECV (PRNTID, SVOLUM, IRBUF)
      CALL PVMFBUFINFO (IRBUF, IBUFSZ, IRMSG, IRTID, IPVMST)
      IF (IPVMST .LT. 0) THEN
        WRITE (6,*) 'BAD PVMBUFINFO RECEIVE MISC - EXIT'
        CALL EXIT (1)
      ENDIF
      IF (XFRTYPE .EQ. USEPVM) THEN
        CALL PVMFUNPACK (BYTE1, L1, INTSIZ, 1, IPVMST)
        CALL PVMFUNPACK (BYTE1, LN, INTSIZ, 1, IPVMST)
        CALL PVMFUNPACK (BYTE1, NODE3, INTSIZ, 1, IPVMST)
        CALL PVMFUNPACK (BYTE1, NODE4, INTSIZ, 1, IPVMST)
         ...
      ELSE
        REWIND (VOLUMFN)
        READ (VOLUMFN) L1,LN,NODE3,NODE4,DUMMY,
     1           ICHECK,
```

```fortran
     1                DVBAR,DVDOT,R123,R124,R134,R234,RBAR,
     2                TAREA,U,VNEW,X1,X2,X3,X4,Y1,Y2,Y3,Y4,
     3                Z1,Z2,Z3,Z4,DVOL,VOL
      ENDIF

      ...

CDIR$ END MASTER, COPY (L1, LN, NODE3, NODE4)

      ...

      CALL XMISC (FROMPARENT, USEPVM)

      ...

      CALL VOLUME_T3D(L1,LN,DVBAR,DVDOT,R123,R124,R134,R234,RBAR,TAREA,
     2                U,VNEW,X1,X2,X3,X4,Y1,Y2,Y3,Y4,Z1,Z2,Z3,Z4,
     3                ICHECK,NODE3,NODE4,DVOL,VOL,lnl1)

      ...

      CALL PVMFINITSEND (PVMDATADEFAULT,ISBUF)
      IF (ISBUF .le. 0) then
         WRITE (6,*) 'Error on initsend in SOLID - exit'
         call exit (1)
      endif
      IF (XFRTYPE .EQ. USEPVM) THEN
         PVMRBUF = DVBAR
         CALL PVMFPACK (REAL8, PVMRBUF, MXLB, 1, IPVMST)
         PVMRBUF = DVDOT
         CALL PVMFPACK (REAL8, PVMRBUF, MXLB, 1, IPVMST)
         PVMRBUF = R123
         CALL PVMFPACK (REAL8, PVMRBUF, MXLB, 1, IPVMST)

         ...

      CALL PVMFSEND (PRNTID, SVOLUMR, IPVMST)

      ...

      RETURN
      END
```

```
       SUBROUTINE VOLUME_T3D(L1,LN,DVBAR,DVDOT,R123,R124,R134,R234,RBAR,
   2              TAREA,U,VNEW,X1,X2,X3,X4,Y1,Y2,Y3,Y4,
   3              Z1,Z2,Z3,Z4,ICHECK,NODE3,NODE4,DVOL,VOL,lnl1)
       ...
c   Shared to private conversion code
       ...
       RETURN
       END
```

Subroutine EPICPV

This subroutine defines a COMMON statement for all the variables used in the various files for data message passing using PVM. It defines the values for various parameters used in PVM.

RESULTS AND DISCUSSIONS

Examples from Research EPIC94 is used to test this new PVM EPIC. The First PVM EPIC run subroutines VOLUME and SOLID in T3D while run the rest of EPIC in YMP. The second PVM EPIC run subroutines VOLUME and SOLID in T3D and the rest of the EPIC is also run in one PE of T3D.

(1)   The first PVM EPIC code

Examples x1 and x4 from Research EPIC94 is used here to compare the results between YMP EPIC code and PVM EPIC code.

YMP code for x1

CPU: 1.06 sec.

Data transferred: 0.848 Mwords.

I/O requests: 92.

PVM EPIC code for x1

	Array size 128	Array size 256
System CPU	89 sec.	63 sec.
Users CPU	33 sec.	26 sec.
Data transferred	23.8 Mwords	28.9 Mwords
I/O requests	5.1E04	3.5E04

YMP code for x4

CPU: 9 Min.

PVM EPIC code for x1

	Array size 128	Array size 512
System CPU	345 Min.	105 Min.
Users CPU	109 Min.	37 Min.
Data transferred	4074 Mwords	4086 Mwords
I/O requests	12.3E06	3.15E06

The CPU time for this case is much larger than the YMP code due to the large number of data transferred and I/O request. The CPU time can be drastically reduced when the number of data requested and I/O requests are reduced. We had used one I/O requests to transfer a cluster of arrays instead of one I/O request for each array. The system CPU is reduced from 105 Min to 9.7 Min and users CPU is reduced from 37 Min to 7.9 Min for example x4. It indicates that the first PVM EPIC code can have good performance if more subroutines are converted to massively parallel program.

(2)   The second PVM EPIC code

Example x4 from Research EPIC94 is used here to compare the results between YMP EPIC code and PVM EPIC code.

YMP code

CPU: 9 Min.

PVM EPIC code

Number od PE	CPU (Min.)
1	42
2	34
4	31
32	28

It was found from the execution that 40% of the example x4 is run in parallel in this PVM EPIC code. From the CPU time shown in the above table, it also shows that 40% of the program is paralleled. Therefore, if 80% of the EPIC code can be paralleled, this PVM EPIC code will run as fast as EPIC YMP when 32 PEs are used. If 90% of the EPIC code can be paralleled, this PVM EPIC will double the speed of EPIC YMP when 32 PEs are used.

CONCLUSION

This study shows that both version of PVM EPIC codes are successfully run in the T3D using PVM. The results show that the speed of both versions can pass the YMP code if more than 80% of EPIC code can be paralleled. Which version is better? This question can not be answered until the EPIC code is completely paralleled. Fortunately, if one version is completely paralleled, the

other version can be done with little extra works.

ACKNOWLEDGEMENTS

ORGANIZED SUPRAMOLECULAR STRUCTURES IN ORGANIC AND POLYMERIC MATERIALS

DR. VLADIMIR V. TSUKRUK
Associate Professor
COLLEGE OF ENGINEERING AND APPLIED SCIENCES

WESTERN MICHIGAN UNIVERSITY
KALAMAZOO, MI 49008

Final Report for:
Summer Faculty Research Program
Wright Laboratory

September, 1995

ORGANIZED SUPRAMOLECULAR STRUCTURES IN ORGANIC AND POLYMERIC MATERIALS

DR. VLADIMIR V. TSUKRUK

Associate Professor

COLLEGE OF ENGINEERING AND APPLIED SCIENCES

WESTERN MICHIGAN UNIVERSITY

Abstract

In the present work we explore organized molecular films of various polymeric materials, dye molecules, and NLO materials built by self-assembling electrostatic deposition and cyclic liquid crystalline (LC) compounds with mesogenic groups attached to cyclic siloxane rings.

It has been demonstrated that a relatively thick (up to 200 nm), rather smooth and homogeneous self-assembled films for copper phthalocyanine dye molecules and synthetic polypeptide, polylysine, can be fabricated by electrostatic layer-by-layer deposition. Microroughness of these films does not exceed 3 nm and a bilayer of 2 nm thick is formed by flexible polylysine fragments and phthalocyanine molecules. Kinetics of formation of self-assembled monolayers was monitored for polystyrene sulfonate (PSS) and polyallylamine (PAA) polymers adsorbed on charged surfaces. A gradual coverage of the surface by isolated polymer islands was observed within the time interval of 5 minutes followed by a formation of an incomplete monolayer film. The processes observed is reminiscence of growth controlled by diffusion limited aggregation.

Computer simulation of cyclic LC compounds reveal some features of their structural behavior in the mesomorphic state. Analysis of X-ray data shows that a variety of scattering phenomena can be explained by overlapping of modulated molecular form-factor with sharper singularities due to association of several molecules along the c-axis ("strings"), presence of siloxane rings with very high scattering power, and lattice factor with body-centered symmetry. A lattice factor features very small lattice sizes (L = 20-30Å) and large distortion (g = 10%) along the a and b-axes and limited correlations along the c-axis: L = 100 - 300Å and g = 2 - 4%. The attachment of mesogenic groups to the ring breaks a symmetry of molecules that results in systematic extinction of odd orders of Bragg reflexes along the c-direction (001, 003, 005 ...) observed for all cyclic compounds studied.

ORGANIZED SUPRAMOLECULAR STRUCTURES IN ORGANIC AND POLYMERIC MATERIALS

DR. VLADIMIR V. TSUKRUK

Introduction.

Supramolecular assemblies and mesomorphic organic molecules belong to a very intriguing class of the matter named "soft matter". [1] These functional polymeric materials able to self-organize in the bulk state are currently of special interest in the field of modern molecular engineering. The ability to form equilibrium phases intermediate between a crystalline state and an isotropic fluid through self-organization offers a promising route toward supramolecular organizations with a high level of complexity and flexibility. A major current trend is the focus on intermolecular interactions that induce suitable supramolecular organizations and provide the possibility for a fine tuning of macroscopic physical (optical, conductive, mechanical) properties. Non-linear optical (NLO) polymers, liquid crystalline (LC) materials with reorientation of molecular fragments in electric fields, and supramolecules with switchable conformations can be thought as prospective materials.

Currently, several approaches exist for fabrication of organized molecular assemblies from functional "soft" materials. [2] The Langmuir-Blodgett (LB) technique is suitable for the fabrication of organized monolayers from amphiphilic molecules at the air-water interface and their subsequent multiple transfer onto a solid substrate. Layer-by-layer or electrostatic deposition exploits Coulombic interactions between oppositely charged molecules physically adsorbed from dilute solutions in an alternative route. Chemisorption of molecules with functional terminal groups like alkylsilanes on the solid substrates with the "right" surface groups results in the formation of robust uniform monolayers chemically tethered to this surface. It has been shown that in the framework of the molecular engineering approach, rather perfect multilayer films (hundreds of layers) can be built with various combinations of molecular fragments, organic and inorganic layers, molecules with switchable conformation, supercells, inhomogeneous distribution of electron density, etc. [2-5]

Goal and objectives.
The goal of this work is twofold. First, we explored ways towards building organized molecular films of various polymeric materials, dye molecules, and NLO materials. We studied monolayer

and multilayer films built of charged polymer (polyelectrolyte), NLO organic dyes, and charged latexes on self-assembled monolayers chemisorbed on glass. Two key elements of the self-assembling process are common for all systems and require special attention: stability of the films that is pre-determined by the level of tethering of the first molecular layer to a solid substrate and concentration of defects that depends upon homogeneity, affinity, and microroughness of this first monolayer.

Second, we studied several LC compounds that show properties promising for their prospective usage as a component of organized supramolecular films. Cyclic LC compounds are composed of mesogenic groups attached to siloxane rings of different sizes.

Experimental

Materials.

We studied functional surface of glass modified by amine-silanes, kinetics of self-assembling of a monolayer film of polystyrene sulfonate (PSS) and polyallylamine (PAA) (Gelest and Aldrich) and multilayer films of copper phthalocyanine tetrasulfonate (CuPc)/amidine PS latex, CuPc/polylysine (Ply), and CuPc/Alcian Blue (AB) dye. Chemical formulae of the components designed to build multilayer films are demonstrated in Figure 1. These films combine a NLO dye molecule, CuPc, with various counter polyions or charged dyes

Figure 1. Schemes of some macromolecular compounds studied

(see Ref. 6 for discussion of optical properties of these films and details of their fabrication procedures). The PS latexes used are highly monodispersed with a diameter of 15 nm and standard deviation of diameter distribution below 5%.

Cyclic LC compounds studied were cyclic siloxanes with biphenyl-containing mesogenic groups (for chemical formulae see Figure 2).

Experimental procedures

Various self-assembled monolayers and multilayers were fabricated on silicon and glass surfaces and studied by combined SPM and X-ray reflectivity.

Film fabrication. An electrostatic layer-by-layer deposition technique was employed for the formation of the films. Solid substrates used were silicon wafers and float glass. Cleaning and modification of the surfaces as well as formation and transfer of the monolayers onto the solid supports was performed using rigorous procedures.[2] All films were made by Dr. A. Campbell in a clean room, class 100, WL (see Ref. 6 for detailed description).
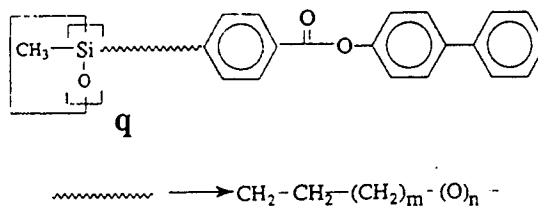
• I, q=5, m=1, n=1 : i 175 n 112 i

• II, q=4, m=1, n=1; i 200 n 180 k

• III, q=5, m=3, n=1; i 172 n 140 Sc 120 k

Figure 2. Chemical formulae of cyclic LCs.

Scanning Probe Microscopy. Surface characterization of molecular films was accomplished by means of scanning probe microscopy (SPM).[7] A key principle of the SPM technique is the probing of a solid surface with a tiny, atomically sharp tip interacting with local groups of atoms. Forces on and displacements of the SPM tip are monitored by highly sensitive electronic feedback and a very precise piezoelement. Atomic force (AFM) and friction force (FFM) images in contact and non-contact (the "tapping") modes were obtained at ambient temperature with the Nanoscope IIIA - Dimension 3000 (Digital Instruments, Inc., 1995) according to the well-established procedures.[7] Scanning was made on scales from 200 nm to 100 μm with the normal load in the range of 1 - 400 nN. All SPM measurements were done by Dr. V. Bliznyuk at WMU.

X-ray reflectivity. From the analysis of X-ray reflectivity the average thickness of molecular films, roughness averaged over a macroscopic scale, and the density distribution of the molecular films can be established.[8] X-ray reflectivity measurements were performed over the range of scattering angles $0 < 2\Theta < 9°$ with steps in the range of $0.005°$ to $0.02°$ on a Philips-MRD instrument (CuKα, $\lambda = 0.154$ nm). We used two different set-ups: four-crystal pinhole collimation for high-resolution measurements (resolution is up to 4000Å) and low-resolution/high-intensity slit collimation (resolution is about 450Å).

X-ray data for cyclic LCs were obtained by Dr. T. Bunning at CHESS, Cornell University.

60-5

<u>Computer modeling.</u> Molecular modeling was performed on a Silicon Graphics Power series workstation using a CERIUS2 computer simulation package. [9] We used energy minimization and molecular dynamics to build an isolated molecule with the lowest energy and a crystal lattice minimization and orthorhombic primitive and centered unit cell to pack the molecules in crystal lattice. To evaluate molecular form-factors, we used pair correlation functions along with the Debye formula to calculate isotropic and meridian scattering. We simulated one-dimensional scattering ("powder diffraction") and two-dimensional pattern ("fiber diffraction"). We took into account different finite sizes of scattering regions and level of distortions in various directions as well as orientational ordering of molecular fragments and thermal fluctuations. All these parameters were derived from experimental data according to the well-known approaches. [10]

Results and discussion

Self-assembled multilayer films.

We observed that the amine-silane monolayer on the glass surface not only provided the positive surface charges necessary to initiate self-assembly but substantially reduced microroughness of the glass substrate: from 2 - 3 nm within 1 μm x 1 μm area to 0.5 - 1 nm. However, island microstructure of the monolayer with submicron lateral dimensions was observed for various preparation conditions and the nature of the substrate (Figure 3a). Obviously, such inhomogeneities can be responsible for micron defects in multilayer self-assembled films as demonstrated for CuPc/AB five bilayer films (Figure 3b). Observed dendrite-like defects propagate through the total thickness of the films and can be caused, for example, by non-uniform drying process. A general tendency for all films studied was the formation of a relatively rough and inhomogeneous surface morphology for the first 1-10 bilayers deposited. Fluctuation of thickness for these films was within 4-25%. However, much smoother surface was observed for thicker films with a number of bilayers up to 100. For these films, the rms microroughness was on a level of 1 - 4 nm (except latex based films with the rms microroughness of about 40 nm) which constitutes only 0.5 - 3% of total thickness.

Negatively charged PSS polymer and PS latex can be adsorbed at the positively charged surface of amine-silane SAM to form a stable monolayer. We observed a gradual coverage of the surface by polymer islands within the time intervals of 1 second to 5 minutes. After 5 minutes of adsorption time we observed incomplete polymer films (Figure 3c). Complete coverage and the formation of a homogeneous polymer layer of about 2 nm thick and <1 nm microroughness was observed after 20-30 minutes of self-assembly.
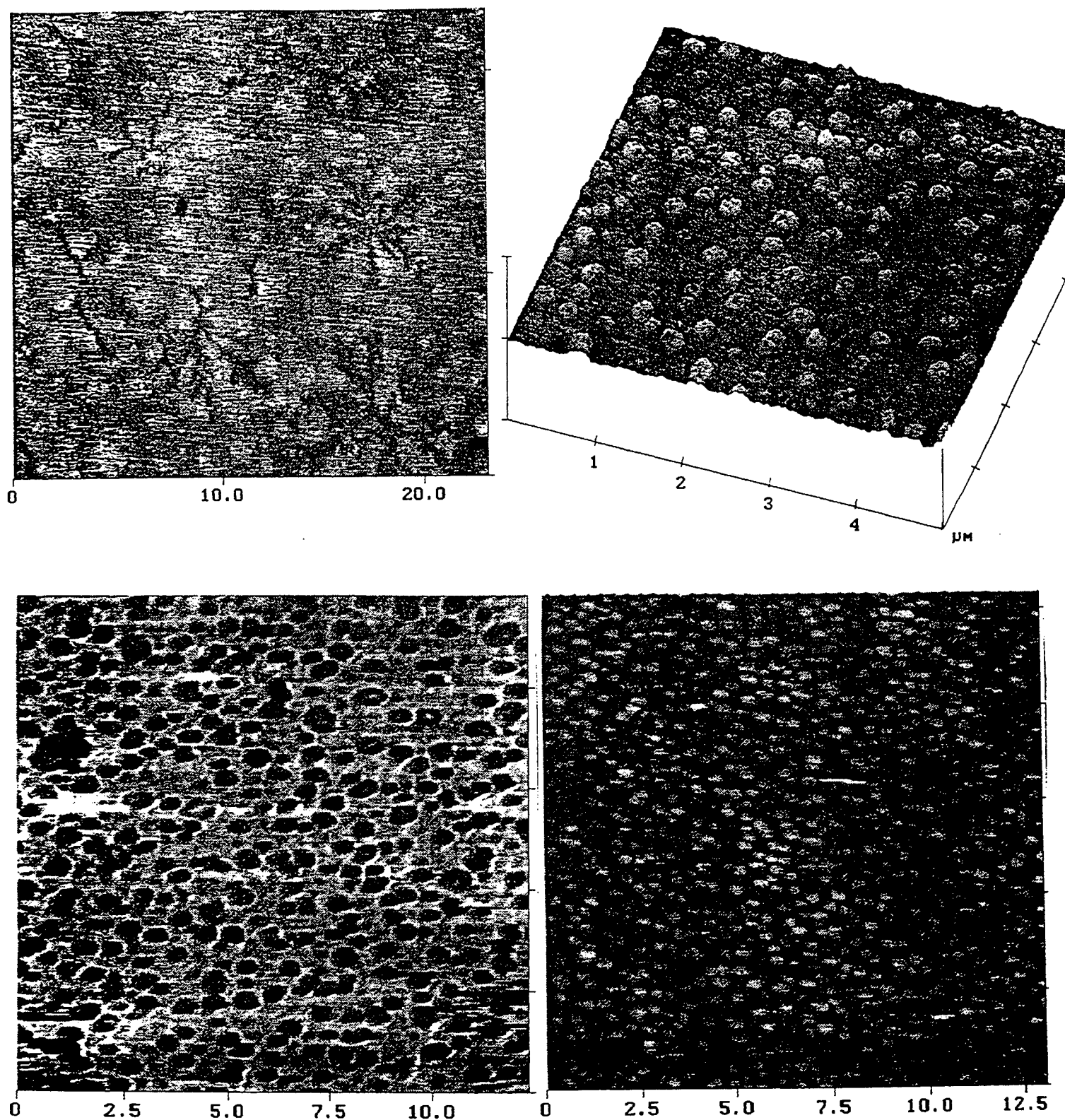
Figure 3. The AFM images of self-assembled monolayers and films (all sizes are in μm):
a) a glass surface coated by amine-silane monolayer with the microroughness of about 1 nm, note the island microstructure of the monolayer; b) a surface of CuPc / Alcian Blue dye film, five bilayers, note dendritic defects propagating through total thickness of the film; c, d) examples of the kinetics of monolayer formation for PSS, 15 minutes (c) and for PS latex, 5 minutes.

For multilayer CuPc/Ply films with the number of bilayers from 5 to 100, we observed a number of Kiessig fringes on X-ray reflectivity curves (Figure 4a). All reflectivity curves show a rapid decay which falls off faster than q^{-4} ($q = 2\sin\Theta/\lambda$). The rapidly decaying reflectivity is modulated by damped oscillations originating from the presence of the thin polymer films with varying degrees of roughness in the air-polymer interface. The periodicity of these modulations depends upon the thickness of the organic layer normal to the surface.
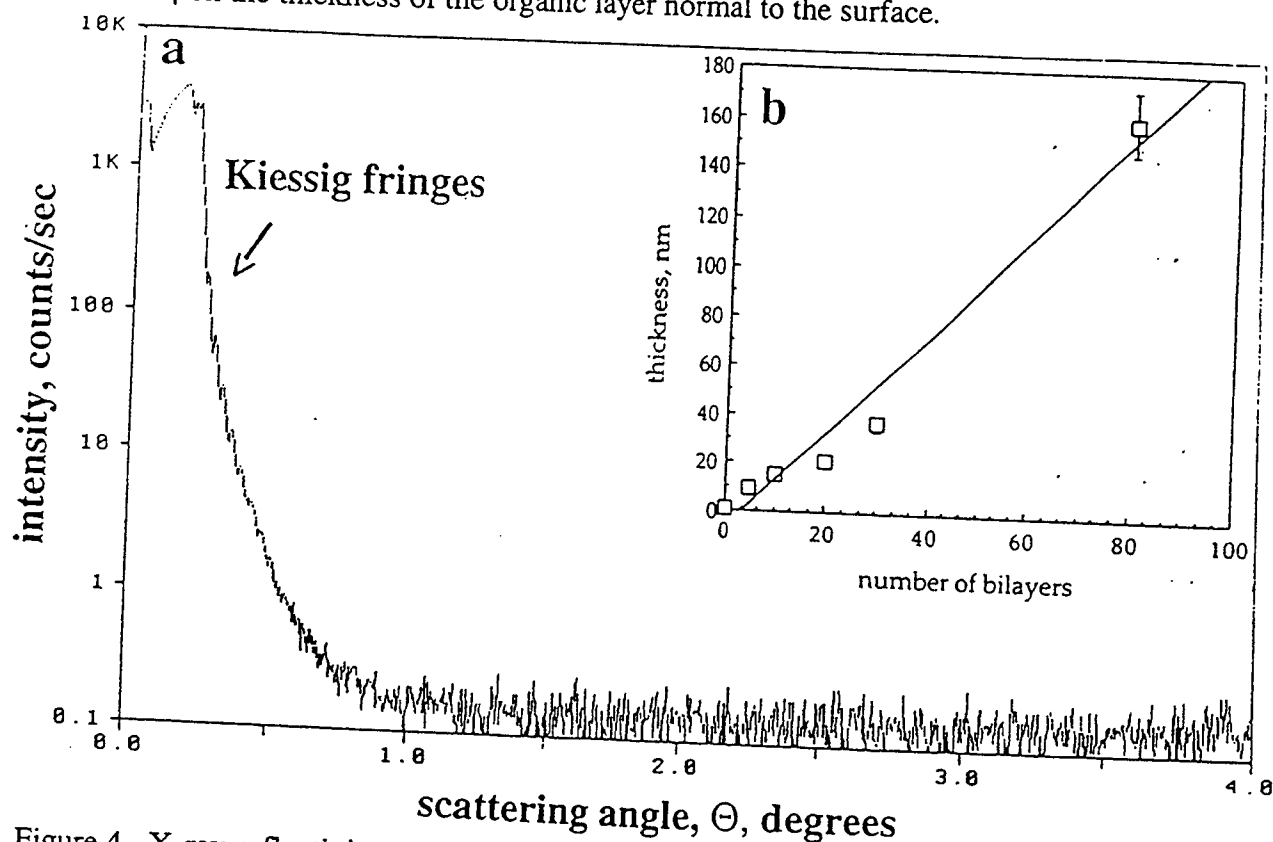
Figure 4. X-ray reflectivity curve for CuPc/Ply film with 80 bilayers and 160 nm thickness (a) and thickness of the films versus number of deposited bilayers (a slope of the fitting line corresponds to average thickness of the bilayer (2.0 nm).

The variation of the film thicknesses with the number of bilayers was estimated from periodicity of the Kiessig fringes for a set of films CuPc/Ply and is shown in Figure 4b. Estimated periodicity of the bilayer structure is 2 nm. However, no Bragg's reflections with periodicity of 2 nm were observed for all films studied. This indicates a very homogeneous density distribution within organized films along the surface normal and, probably, substantial overlapping of adjacent monolayers of CuPc and Ply. The films with a limited number of bilayers (5-20) are relatively rough with microroughnesses estimated from the AFM images within 3 μm x 3 μm areas in the range of 0.8 - 1.2 nm or variation of local thickness up to 10% of an average. Average thickness of the bilayer for these thin films was well below the average periodicity for

thick films (1.1 nm versus 2 nm) due to the nonuniform formation of the films with a limited number of molecular layers. This conclusion is confirmed by ellipsometric measurements. [6]

The CuPc/Ply films with greater than 30 bilayers are relatively smooth and homogeneous as can be judged from the number of the Kiessig fringes observed on the X-ray reflectivity curves (Figure 4a). The AFM observations show relatively uniform surfaces with a low microroughness of 2.7 nm or 1.3% of total thickness for the film with 100 bilayers. The thickness of one bilayer derived from Figure 4b is 2.0 ± 0.1 nm. This value is close to the expected thickness of a bilayer formed by phthalocyanine cores in flat-on position, parallel to the surface plane (thickness is about 0.3 - 0.5 nm) and 1.5 - 1.7 nm thick stratum of polylysine fragments. This thickness s close to one obtained in ellipsometric measurements of the PSS monolayers [11] and to our own AFM observation of a monolayer film of PSS (about 2 nm).

Cyclic LC compounds

Chemical formulas of cyclic LC compounds considered here are shown in Figure 2. As has been shown before, cyclic compounds with siloxane ring and various mesogenic groups display some unusual phenomena and structural behavior in electric fields (see Refs. 12). One of them is the observation of multiple diffuse but rather strong small-angle dash-like reflexes in the oriented nematic phase. Up to six reflexes with modulated intensity is usually observed as demonstrated in Figure 5 for compound III. Several possible theories were proposed to explain this phenomena. String formation and correlated displacements of molecular rows are possible explanations. [13] For selected LC compounds which possess similar behavior we studied the role of intramolecular interference (molecular form-factor), combination of intra- and intermolecular contributions, and a lattice factor by means of computer simulation.
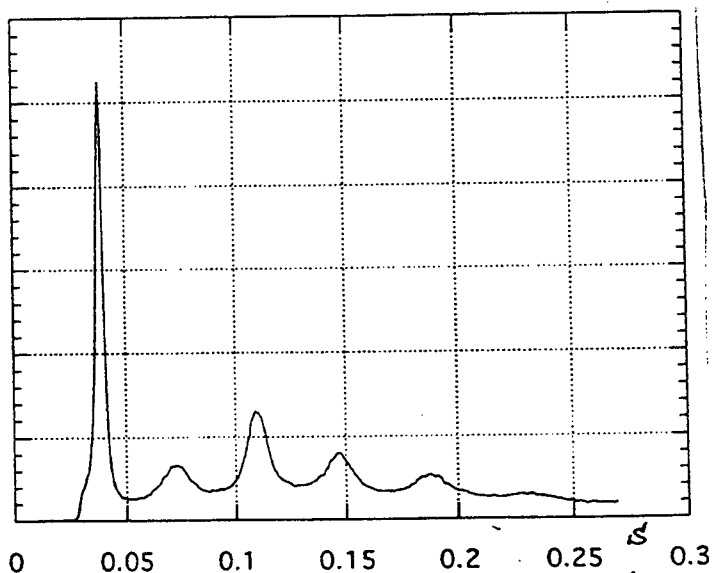
Figure 5. X-ray scattering for compound III in the nematic phase.

An example of a model for a single molecule built and used for X-ray simulations is presented in

Figure 6a. This and other computer models were built by following procedure:

- an initial model was built using a Builder routine; - five cycles of molecular dynamics (anneal dynamics) with minimization step between cycles followed by "deep" minimization (500 steps) were applied to find molecular conformation with the lowest total energy; - then the molecule was minimized under anisotropic external pressure: compression along the x and y axes and tensile along the z-axis. Pressure applied was in the range of 0.1 to 30 kbar. The highest pressure allows the molecular fragments to change their conformation to adopt dense packing with closest intermolecular contacts; - the molecule was placed within a crystal lattice and volume of the unit cell was minimized (under modest external pressure) with cycles of energy minimization of the molecule in-between; - the resulting molecule was a subject of "relaxation procedure" that includes several cycles of molecular dynamics and energy minimization without external pressure; - finally, the molecule was placed within the unit cells of different symmetry and energy of packing was minimized under modest pressure keeping the molecule as a rigid unit and allowing variations of cell parameters in combination with translation and rotation of the molecular unit as a whole.
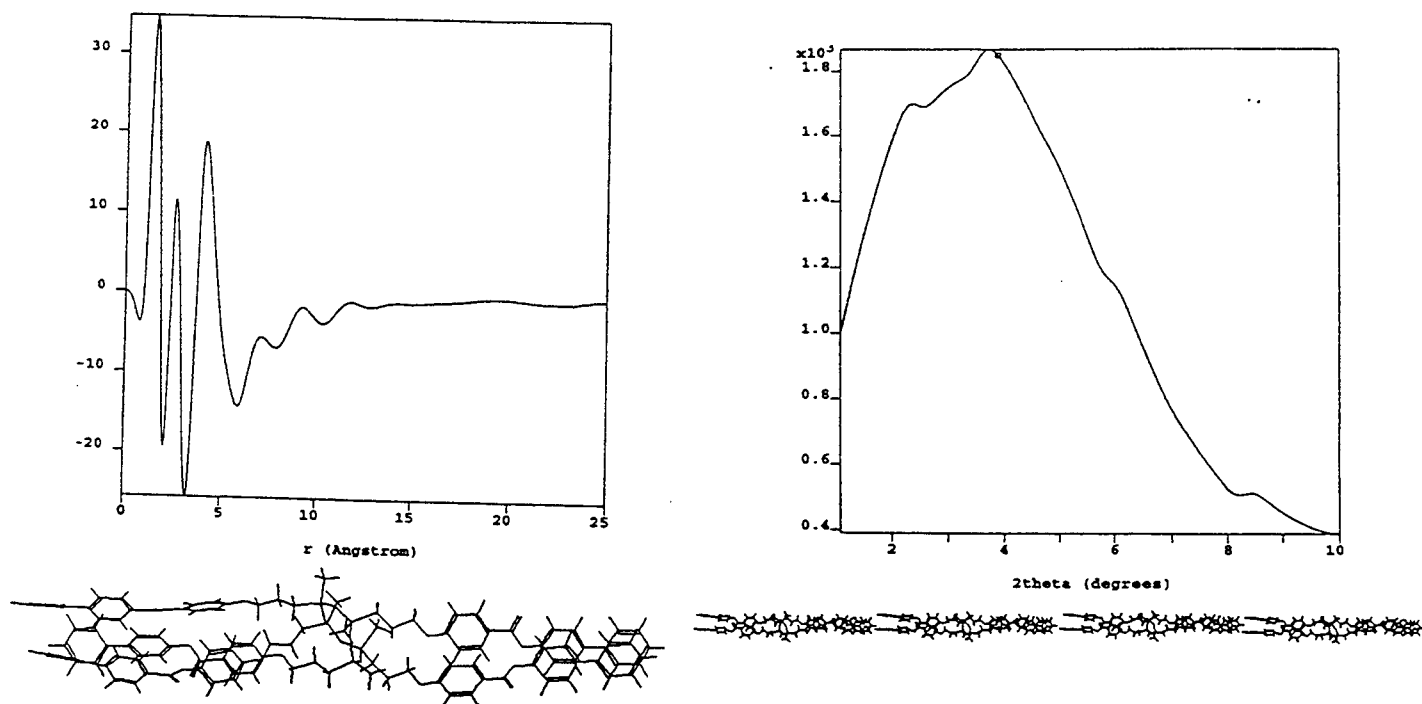
Figure 6. Molecular models of compound III and pair correlation function (left) and a model of "string" and simulated isotropic X-ray scattering.

To evaluate structural properties of the molecules we calculated pair correlation functions (or radial distribution functions, RDF) and X-ray scattering (isotropic, one-dimensional, and two-

Cylindrical Scattering
Radiation used = XRAY
Reduced scattering
i(R,Z)
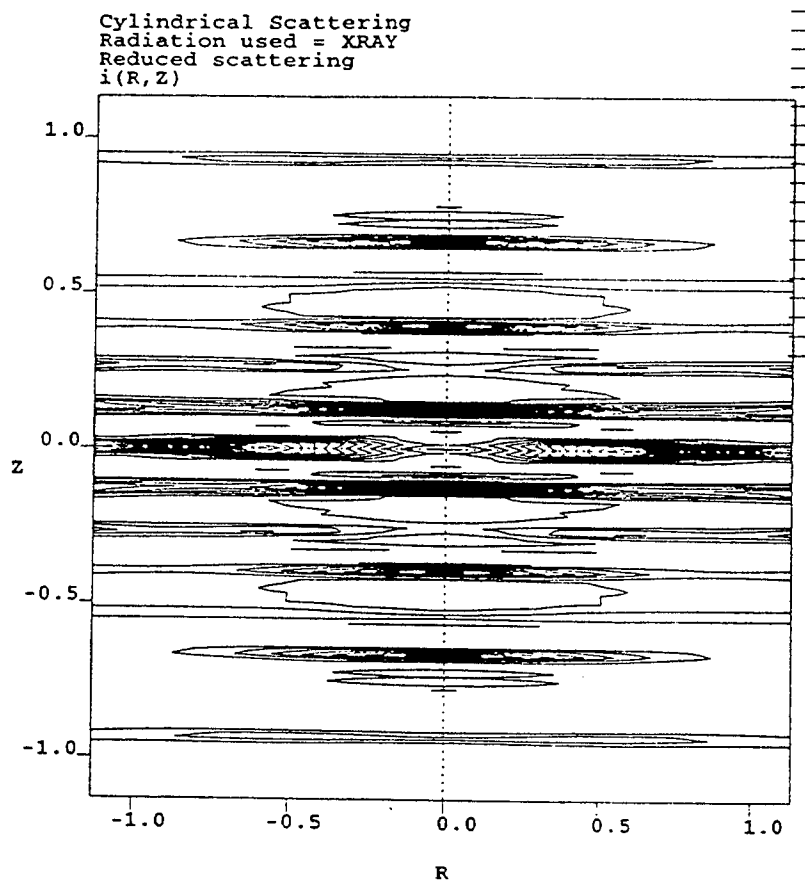
Fiber Diffraction
Radiation used = XRAY
Wavelength   =   1.5418
Tilt angle   =   0.0000
monoc8004   FLAT-PLATE
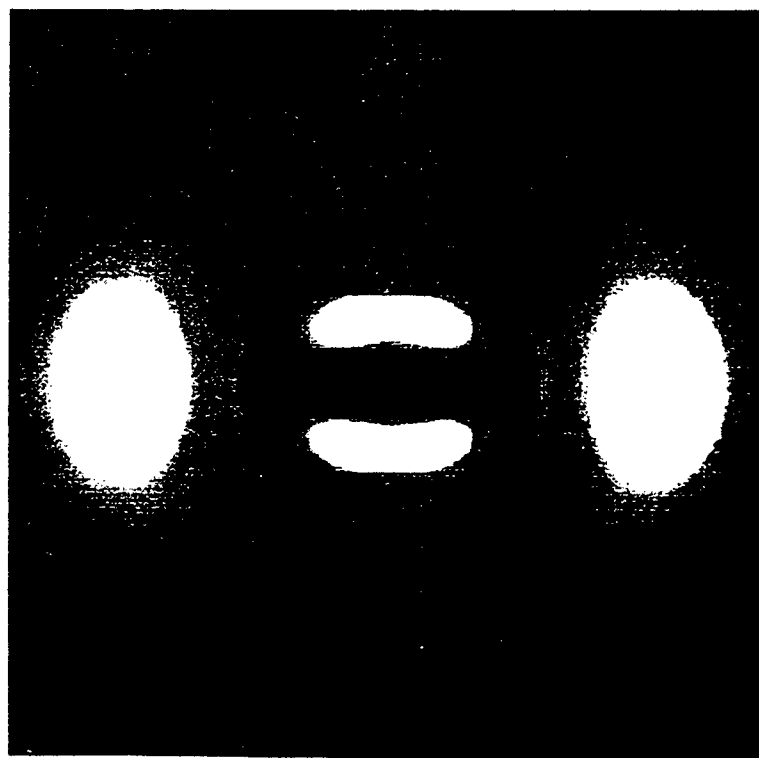
Figure 7. Simulated cylindrical scattering for a "string" model (top right) of compound III (top) and simulated fiber pattern for a crystal lattice of a single mesogenic group (bottom)

dimensional) by a single molecule. X-ray scattering was simulated for associations of two to eight molecules aligned in a single "string" or associated strings with correlated displacement along the molecular axes. Finally, the total scattering function for a crystal lattice with limited sizes in various directions and a variable level of distortions was calculated.

As an example, simulated pair correlation functions, isotropic scattering intensity, and cylindrical scattering pattern are presented in Figure 6 and 7a for compound III along with models of a single molecule and a string composed of four molecules. Figure 7b shows an example of fiber diffraction simulated for a model mesogenic group packed in a distorted crystal lattice. Finally, an example of X-ray scattering curve for compound III in the smectic phase in comparison with the simulated [00L] lattice factor for the body-centered orthorhombic unit cell is demonstrated in Figure 8. For this crystal lattice, the length of the c-axis of the unit cell (horizontal direction in Figure 6a) is 57Å and lattice size and distortion factor in c-direction are 400Å and 1%, respectively.

To make firm conclusions about structural ordering in the compounds studied, detailed analysis of simulated results in conjunction with a broad spectrum of experimental X-ray data for different temperatures, parameters of electric field, and architecture of molecular fragments should be completed. Such an analysis is undergoing and the results will be published elsewhere. Here, we can make only some general conclusions.

First, analysis of partial pair correlation functions for various atoms (O-O, Si-O, Si-C) shows that in addition to standard set of atom-atom distances (1.5Å for C-C, C-O and 2.5 Å for C-C-C), additional sets of intramolecular distances of 7Å, 9Å, 11Å, 15Å, 23Å exist along molecular axis. These sets are due to combination of partial contributions of intramolecular distances with high scattering power (for example, the contribution of Si-O = 12 C-C). These contributions are related to the distances between the siloxane ring and the oxygen atoms located in different parts of the mesogenic groups. This statistic of intramolecular distances can explain the presence of diffuse maxima on the form-factor of a single molecule (Figure 6b).

Second, preliminary analysis of X-ray data shows that a variety of scattering phenomena can be explained by overlapping of a modulated molecular form-factor with sharper singularities due to the association of several molecules along the c-axis ("strings") and the lattice factor for the body-centered unit cell. The number of molecules in a "string" should be in the range of three to eight and a substantial level of longitudinal correlations might be present between neighboring

"strings" for some compounds. A lattice factor features very small lattice sizes (L = 20-30 Å) and large distortion (g = 10%) along the a and b-axes of the crystal lattice. More pronounced correlations are observed along the c-axis: L = 100-300 Å and g = 2 - 4% for various compounds. For some phases we have to assume the body-centered triclinic unit cells and significant molecular tilt. To discriminate between various models additional analysis of existed experimental data and additional experiments are required.

Third, we observed that for LC compounds similar to those discussed above, but without a siloxane ring, the most of the observed X-ray scattering phenomena can be easily explained by the formation of a distorted lattice with the primitive unit cell.
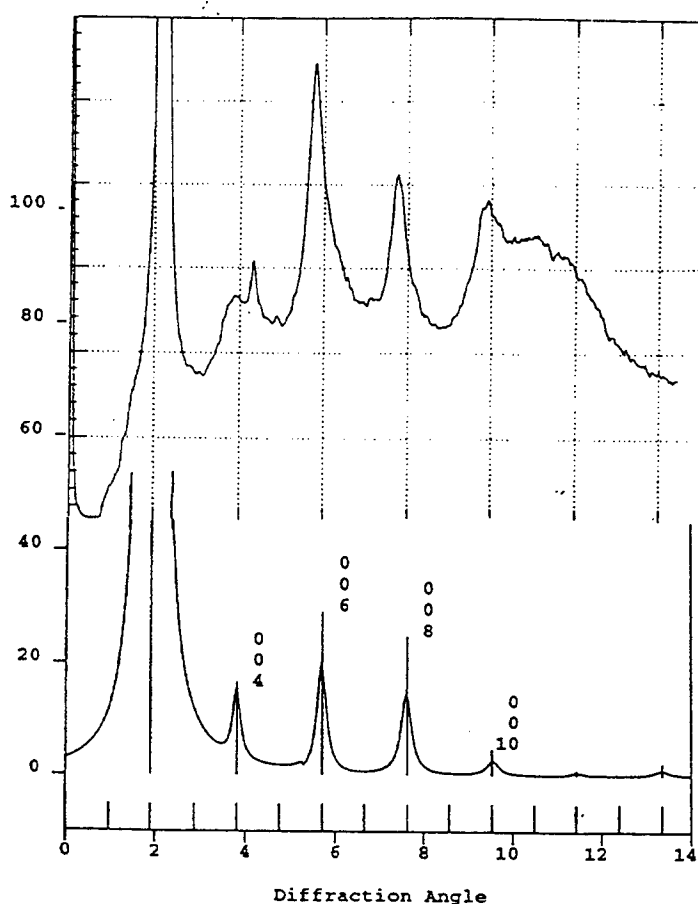
Figure 8. X-ray scattering in the smectic phase of compound III (top) and simulated diffraction for the orthorhombic body-centered unit cell (bottom).

Obviously, the chemical attachment of the mesogenic groups to a siloxane ring with a strong scattering power (equivalent to 40 carbon atoms concentrated within small volume of about 30 $Å^3$) causes significant increase of the contribution of the molecular form factor in total scattering pattern. Moreover, this attachment breaks a symmetry of molecules that results in the systematic extinction of odd orders of Bragg reflexes along the c-axis of the unit cell (001, 003, 005 ...) observed for all cyclic compounds studied.

Conclusions

On the basis of our results we can make several general conclusions about supramolecular structural organization of organic compounds studied.

By electrostatic layer-by-layer deposition a relatively thick (up to 200 nm), rather smooth and homogeneous self-assembled films for dye molecules (CuPc) and synthetic polypeptide can be fabricated. The microroughness of these films does not exceed 3 nm and a bilayer of 2 nm thick is formed by flexible polylysine fragments and phthalocyanine molecules in flat-on position. All films fabricated with a limited number of layers (5 to 20) display uneven surface morphology and incomplete molecular layers. However, the AFM observations show a uniform surface with a low microroughness of 2.7 nm or 1.3% of total thickness for the film with 100 bilayers.

The kinetics of formation of self-assembled monolayer driven by Coulombic interactions was monitored for PSS polymer adsorbed on charged amine-silane SAM. A gradual coverage of the surface by isolated polymer islands was observed within the time interval of 5 minutes. An incomplete monolayer films were observed at longer times of adsorption. The formation of a complete monolayer of about 2 nm thick was observed only after 20-30 minutes of self-assembly. The kinetic processes observed reminisce growth controlled by diffusion limited aggregation.

Computer simulation of cyclic LC compounds reveals some features of their structural behavior in the mesomorphic state. Analysis of X-ray data shows that a variety of scattering phenomena can be explained by overlapping of modulated molecular form-factor with sharper singularities due to the association of several molecules along the c-axis ("strings"), the presence of siloxane rings with very high scattering power, and a lattice factor with body-centered symmetry. The number of molecules in a single "string" should be in the range of three to eight. A substantial level of longitudinal correlations might be present between neighboring "strings" for some LC compounds. The attachment of the mesogenic groups to a siloxane ring breaks a symmetry of molecules that results in the systematic extinction of odd orders of Bragg reflexes along the c-direction (001, 003, 005 ...) observed for all cyclic compounds studied.

Acknowledgments

References

1. de Gennes, P. *Soft Matters*, Nobel Prize Lecture, *Science, 256,* 495, 1992

2. Ulman, A. *"Introduction to Ultrathin Organic Films"*, Acad. Pres., San Diego, 1991

3. Tsukruk, V. V. in: *The Polymeric Materials Encyclopedia,* Ed. J. Salamone, CRC Press, 1995

4. Lvov, Yu. M., Decher, G. *Crystallography Reports, 39,* 628, 1994

5. Fendler, J. H.; Meldrum, F. C. *Adv. Materials, 7,* 607, 1995

6. Cooper, T. M., Campbell, A. L., Crane, R. L. *Langmuir, 11,* 2713, 1995

7. Tsukruk, V. V., Reneker, D. H. *Polymer, 36,* 1791, 1995

8. Foster, M. *Crit. Rev. Anal. Chem., 24,* 179, 1993

9. CERIUS2, Molecular Simulation Corp., Cambridge, UK, 1994

10. Tsukruk, V. V., Shilov, V. V. *Structure of Polymeric Liquid Crystals,* Kiev, 1990

11. Miyano, K., Asano, K., Shimomura, M. *Langmuir, 7,* 444, 1991

12. McNamee, S. G., Bunning, T. J., McHugh, C.M., Ober, C. K., Adams, W. W. *Liquid Crystals, 17,* 179, 1994

13. Davidson, P., Levelut, A. M. *Liquid Cryst., 11,* 469, 1992

ELECTROMECHANICS OF SEGMENTED CYLINDRICAL PIEZOELECTRIC SENSOR/ACTUATOR PATCHES

H. S. Tzou

Department of Mechanical Engineering

University of Kentucky
CRMS–414N
Lexington, KY 40506–0108

Electromechanics of Segmented Cylindrical Piezoelectric Sensor/Actuator Patches

H. S. Tzou[1], Y. Bao[1], and V. B. Venkayya[2]

[1] Department of Mechanical Engineering
University of Kentucky
Lexington, KY 40506–0046

[2] Wright Laboratory
Flight Dynamics Directorate
WL/FIBA, WPAFB Ohio 45433

ABSTRACT

Spatial characteristics, modal sensitivities, modal filtering, curvature effects, etc. of distributed segmented cylindrical sensor/actuator patches are investigated. A sensor equation suggests that the sensor signal is determined by a number of factors, such as geometries, material properties, mode numbers, sensor locations, spatial distributions, strains, etc. The total sensor sensitivity is composed of a membrane sensitivity and a bending sensitivity which are respectively related to induced membrane strains and bending strains. A number of sensor parameters (e.g., sensor thickness, shell thickness, curvature angles, shell sizes) are evaluated, and their membrane, bending, and total sensitivities are compared. Modal control forces of segmented actuator patches are derived and evaluated. Modal actuation factor, modal feedback factor, and controlled damping ratio are derived and their detailed membrane and bending actuations are evaluated with respect to actuator design parameters: actuator thickness, shell lamina thickness, shell curvatures, shell sizes, and natural modes.

Electromechanics of Segmented Cylindrical Piezoelectric Sensor/Actuator Patches

H. S. Tzou, Y. Bao, and V. B. Venkayya

INTRODUCTION

Sensing/control effectiveness and spatial characteristics of distributed piezoelectric sensors and actuators depend on a number of factors, such as material properties, electrode properties, placements (surface bonded or embedded), spatial shaping, spatial distributions, thickness variations, etc (Tzou, 1993; Tzou and Anderson, 1992; Tzou and Fukuda, 1992). Distributed piezoelectric sensors respond to mechanical strains and generate electric signals (either charge or voltage) due to the direct piezoelectric effect. Distributed piezoelectric actuators are induced—strain type actuators in which the control strains are induced by high voltages or charges due to the converse piezoelectric effect. Depending on placements, distributed sensors can sense either membrane responses, bending responses, or both. Distributed actuators can provide membrane control actions and bending control actions. Spatial thickness or surface shaping of distributed sensors/actuators can also make them respond to a vibration mode (Lee, 1992; Gu, et al., 1994; Tzou, Zhong, Natori, 1993) or a group of modes (Hubbard & Burke, 1992; Tzou, 1993). Distributed sensing and control of piezoelectric laminated <u>beams</u> were investigated (Lee, 1992; Collins, et, al., 1994; Tzou & Hollkamp, 1994). Two—dimensional (2D) zero curvature structures (e.g., plates) with distributed sensors and actuators were studied (Gu, Clark, Fuller, Zander, 1994; Sumali and Cudney, 1993; Tzou and Tseng, 1991; Tzou and Fu, 1994; Suleman & Venkayya, 1994; Detwiler, Shen, & Venkayya, 1994), and also rings and cylindrical shells (Qiu & Tani, 1994; Tzou, Zhong, Natori, 1993; Tzou, 1993). 2D segmented sensor/actuator patches laminated on rectangular plates were investigated and their spatial sensing/control effectiveness studied (Sumali & Cudney, 1993; Tzou and Fu, 1994).

In this work, based on the piezoelastic shell lamination theory, spatial sensing and actuation of cylindrical shell sensor/actuator patches are investigated. Distributed sensor patches are studied first, and followed by distributed actuator patches. Detailed micro electromechanics, functionality, spatial effectiveness, and sensitivities of segmented shell sensor patches are investigated. In addition, spatial actuation effects and parametric studies of distributed segmented actuators are investigated. Micro electromechanics of

segmented actuator patches are studied and detailed contributions of membrane and bending control effects are analyzed.

LAMINATED CYLINDRICAL SHELL

It is assumed that a simply supported laminated cylindrical shell (with radius R, length L and curvature angle β^*) is made of three elastic laminae (with thickness h_2, h_3, and h_4) sandwiched between two piezoelectric laminae (with thickness h_1 and h_5), respectively. The bottom (1st) layer serves as a sensor layer and the top (5th) layer serves as an actuator layer, Figure 1. Note that the two piezoelectric laminae are elastically continuous, however, their electric properties are restricted by their respective surface electrodes confined by segmented patches. Design parameters, such as shell thickness, sensor/actuator thickness, curvatures, and sizes, are evaluated.
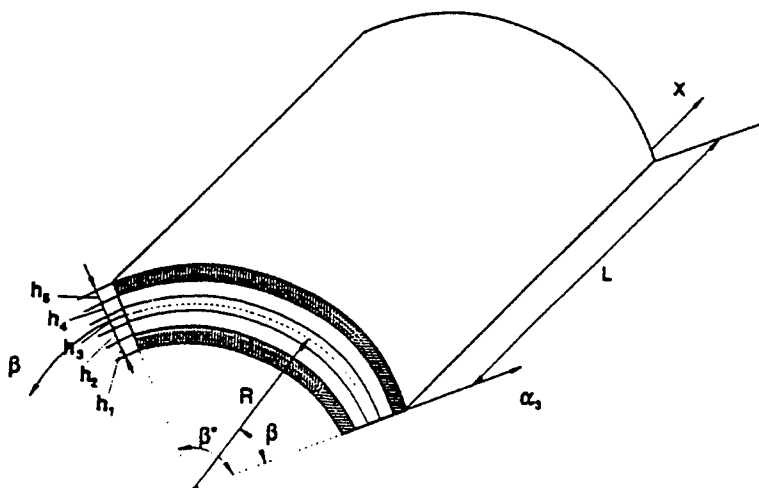
Fig.1 A piezoelectric laminated cylindrical shell.

SEGMENTED SENSOR PATCHES

It is assumed that the sensor layer is completely covered with a uniformly distributed electrode and the electrode is equally segmented into four sensor patches along the center lines. Thus, each sensor patch has an effective area $S_p^e = R(\beta^* L)/4$, Figure 2.

Sensor signal from each sensor patch can be individually calculated based on its spatial distribution defined by its coordinates.
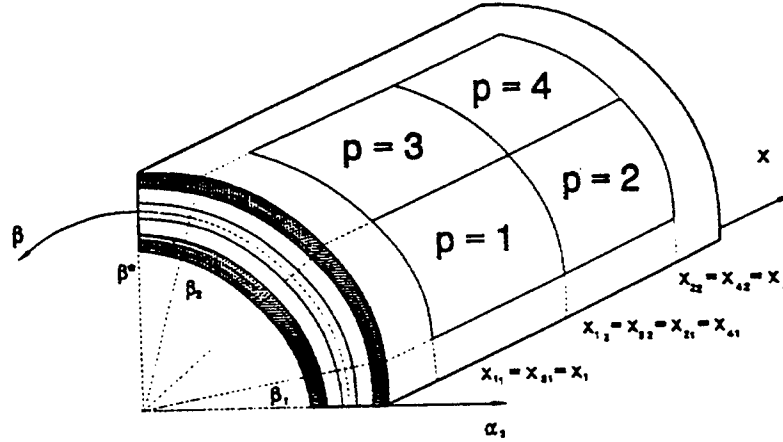
Fig.2 Quarterly segmented piezoelectric patches.

Sensor signals $(\phi_{mn}^{ds})_i$ of the mn—th mode from these sensor patches, numbered in Figure 2, can be respectively calculated by (Tzou, Bao, and Venkayya, 1996)

$$(\phi_{mn}^{ds})_1 = 4\eta_{mn}(t)\, S_{mn}(1 - \cos\tfrac{m\pi}{2})(1 - \cos\tfrac{n\pi}{2}) , \tag{1a}$$

$$(\phi_{mn}^{ds})_2 = 4\eta_{mn}(t)\, S_{mn}(\cos\tfrac{m\pi}{2} - \cos m\pi)(1 - \cos\tfrac{n\pi}{2}) , \tag{1b}$$

$$(\phi_{mn}^{ds})_3 = 4\eta_{mn}(t)\, S_{mn}(1 - \cos\tfrac{m\pi}{2})(\cos\tfrac{n\pi}{2} - \cos n\pi) , \tag{1c}$$

$$(\phi_{mn}^{ds})_4 = 4\eta_{mn}(t)\, S_{mn}(\cos\tfrac{m\pi}{2} - \cos m\pi)(\cos\tfrac{n\pi}{2} - \cos n\pi) . \tag{1d}$$

where η_{mn} is the modal participation factor; S_{mn} is the mn—th *modal sensitivity* denoting the mn—th modal signal per unit modal participation factor [V/m] which is composed of a *membrane sensitivity* $(S_{mn})_{memb}$ and a *bending sensitivity* $(S_{mn})_{bend}$:

$$S_{mn} = (S_{mn})_{memb} + (S_{mn})_{bend} = \frac{h\,^s e_{31}}{mn\epsilon_{33}}\left[\frac{1}{R\pi^2} + r^s\left[(\tfrac{m}{L})^2 + (\tfrac{n}{R\beta^*})^2\right]\right] , \tag{2}$$

where h^s is the sensor thickness; e_{31} is the piezoelectric constant; ϵ_{33} is the dielectric

permittivity; and r^s is a distance measured from the neutral surface to the sensor mid–plane. Note that cosine and one terms in Eqs.(1a–d) are related to sensor's spatial distribution and they are modal dependent. For example,

$$(1 - \cos\frac{n\pi}{2}) = \begin{cases} 1, & n = 1, 3, 5, 7, 9,... \\ 2, & n = 2, 6, 10, 14, ... , \\ 0, & n = 4, 8, 12, 16, ... \end{cases} \qquad (3a)$$

$$(\cos\frac{n\pi}{2} - \cos n\pi) = \begin{cases} 1, & n = 1, 3, 5, 7, 9,... \\ -2, & n = 2, 6, 10, 14, ... \\ 0, & n = 4, 8, 12, 16, ... \end{cases} \qquad (3b)$$

Thus, all quadruples of the m or n modes of the cylindrical shell are not observable. In order to detect these modes, one needs to change the sizes and locations of segmented sensor patches. Detailed numerical analysis of design variables is presented in case studies.

SEGMENTED ACTUATOR AND VIBRATION CONTROL

Recall that the top layer is an actuator layer, the bottom layer is a sensor layer, and the other three middle layers are elastic laminae in the five–layer laminated cylindrical shell configuration. When segmenting distributed sensor and actuator layers into patches, both sensor and actuator layers are segmented into the same patterns such that each actuator patch has a corresponding sensor patch with the same coordinates in x and β. Each sensor signal is processed and fed back to the corresponding collocated distributed actuator patch. Figure 3 illustrates three feedback control schemes of the laminated cylindrical shell (Tzou, Bao, and Venkayya, 1996).
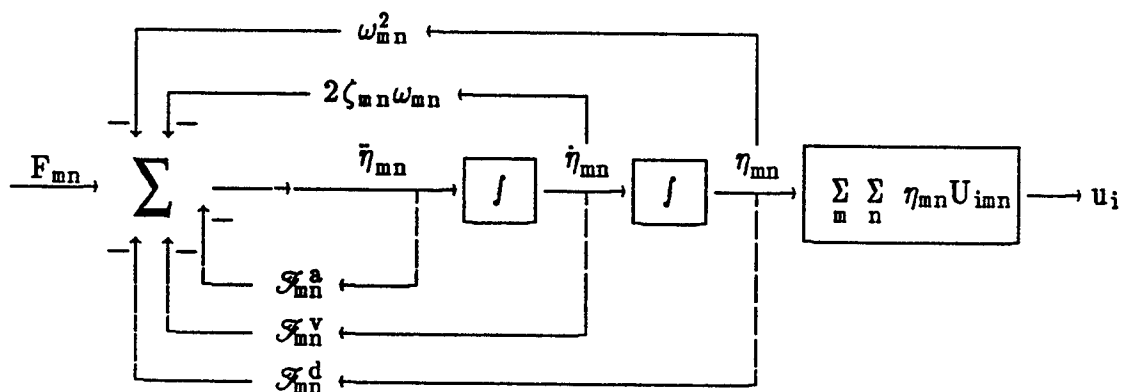
Fig.3 Feedback control of a laminated cylindrical shell.

It is assumed that the distributed piezoelectric actuator is equally segmented into four equal–sized actuator patches, the same pattern as the sensor patches. Considering the modal control force F_{mn}^{c}, one can write the mn–th modal equation as

$$\ddot{\eta}_{mn} + 2\zeta_{mn}\omega_{mn}\dot{\eta}_{mn} + \omega_{mn}^{2}\eta_{mn} = F_{mn}^{m}(t) + F_{mn}^{t}(t) + \sum_{p=1}^{4}(F_{mn}^{c})_{p}$$

$$= F_{mn}^{m}(t) + F_{mn}^{t}(t) - A_{mn}^{a}\left[(\phi^{c})_{1}\cdot(1-\cos\frac{n\pi}{2})(1-\cos\frac{m\pi}{2})\right.$$

$$+ (\phi^{c})_{2}\cdot(1-\cos\frac{n\pi}{2})(\cos\frac{m\pi}{2}-\cos m\pi) + (\phi^{c})_{3}\cdot(\cos\frac{n\pi}{2}-\cos n\pi)(1-\cos\frac{m\pi}{2})$$

$$\left. + (\phi^{c})_{4}\cdot(\cos\frac{n\pi}{2}-\cos n\pi)(\cos\frac{m\pi}{2}-\cos m\pi)\right], \tag{4}$$

where F_{mn}^{m} is the mechanical excitation; F_{mn}^{t} is the temperature excitation; $(\phi^{c})_{i}$ is the control voltage fed into the i–th actuator patch and i = 1,2,3,4 and A_{mn}^{a} is the *modal actuation factor* which can be further divided into a *membrane actuation factor* $(A_{mn}^{a})_{memb}$ and a *bending actuation factor* $(A_{mn}^{a})_{bend}$. Similar to the quarterly segmented sensors, all quadruples of the m or n modes can not be controlled by the quarterly segmented actuators. (To control these modes, other segmentation or shaping techniques need to be implemented.) Positive or negative signs of modal voltages resulting from sensor patches depend on their respective patch locations and mode numbers m and n. Accordingly, signs of feedback voltages to actuator patches need to be carefully monitored. It is also assumed that the temperature excitation frequency differs from shell natural frequencies, and modal filters are used such that each modal equation can be considered independently, i.e., neither observation nor control spillover. Three generic feedback algorithms (i.e., displacement, velocity, and acceleration) and their modal feedback factors are presented in Figure 5. (Note that combination of the displacement and velocity feedback yields the conventional proportional plus derivative feedback.) However, only the velocity feedback is presented, due to page limitations. In the velocity feedback (derivative feedback) control, the *p*–th actuator voltage is proportional to the derivative of the collocated sensor signal. Imposing modal filers to isolate the mn–modal signal, one can write the modal feedback voltage as

$$(\phi^{c})_{p} = \mathcal{G}_{v}(\dot{\phi}_{mn}^{ds})_{p}. \tag{5}$$

Thus, the resultant modal control force from all k actuator patches is

$$F_{mn}^c(t) = -A_{mn}^a \sum_{p=1}^{k} \mathcal{G}_v (\dot{\phi}_{mn}^{ds})_p \left(\cos\frac{m\pi x_{p1}}{L} - \cos\frac{m\pi x_{p2}}{L}\right)\left(\cos\frac{n\pi\beta_{p1}}{\beta*} - \cos\frac{n\pi\beta_{p2}}{\beta*}\right)$$

$$= -\dot{\eta}_{mn}(t) A_{mn}^a S_{mn}^S \mathcal{G}_v \sum_{p=1}^{k} \frac{R\beta*L}{S_p^e}\left(\cos\frac{m\pi x_{p1}}{L} - \cos\frac{m\pi x_{p2}}{L}\right)^2\left(\cos\frac{n\pi\beta_{p1}}{\beta*} - \cos\frac{n\pi\beta_{p2}}{\beta*}\right)^2$$

$$= -\dot{\eta}_{mn}(t) A_{mn}^a S_{mn}^S \mathcal{G}_v \Delta_{mnk} , \tag{6}$$

where $\mathcal{G}_v$ is the velocity feedback gain. The mn–th modal equation becomes

$$\ddot{\eta}_{mn} + (2\zeta_{mn}\omega_{mn} + \mathcal{H}_{mn}^v)\dot{\eta}_{mn} + \omega_{mn}^2\eta_{mn} = F_{mn}^m(t) + F_{mn}^t(t) , \tag{7}$$

where $\mathcal{H}_{mn}^v$ is the mn–th *modal velocity feedback factor* and $\mathcal{H}_{mn}^v = A_{mn}^a S_{mn}^S \mathcal{G}_v \Delta_{mnk}$. For the four equal–sized segmented actuator, i.e., k=4, the spatial effect Δ_{mn4} is

$$\Delta_{mn4} = 4\left[(1-\cos\frac{n\pi}{2})^2(1-\cos\frac{m\pi}{2})^2 + (1-\cos\frac{n\pi}{2})^2(\cos\frac{m\pi}{2}-\cos m\pi)^2\right.$$
$$\left. + (\cos\frac{n\pi}{2}-\cos n\pi)^2(1-\cos\frac{m\pi}{2})^2 + (\cos\frac{n\pi}{2}-\cos n\pi)^2(\cos\frac{m\pi}{2}-\cos m\pi)^2\right] . \tag{8}$$

The modal velocity feedback factor $\mathcal{H}_{mn}^v$ including all four feedback possibilities is

$$\mathcal{H}_{mn}^v = [(F_{mn})_{m,m} + (F_{mn})_{b,b} + (F_{mn})_{m,b} + (F_{mn})_{b,m}]\,\mathcal{G}_v\Delta_{mn4}$$
$$= (\mathcal{H}_{mn}^v)_{m,m} + (\mathcal{H}_{mn}^v)_{b,b} + (\mathcal{H}_{mn}^v)_{m,b} + (\mathcal{H}_{mn}^v)_{b,m} . \tag{9}$$

The subscripts m,m, b,b, m,b, and b,m are defined as follows. The first letter denotes the actuator component effect and the second letter denotes the sensor component effect; m is for the membrane effect and b is for the bending effect. There are membrane and bending contributions to the sensor signals in sensor patches, and also membrane control force and control moment effects in actuator patches. The controlled modal damping ratio ζ_{mn}' includes the original damping ratio ζ_{mn} and the modal velocity feedback factor $\mathcal{H}_{mn}^v$:

$$\zeta_{mn}' = \zeta_{mn} + \mathcal{H}_{mn}^v/2\omega_{mn} , \quad \text{and} \tag{10}$$

$$\mathcal{H}_{mn}^v/2\omega_{mn} = (\zeta_{mn}^v)_{m,m} + (\zeta_{mn}^v)_{b,b} + (\zeta_{mn}^v)_{m,b} + (\zeta_{mn}^v)_{b,m} . \tag{11}$$

Detailed membrane and bending controlled damping ratios are defined by

i) $(\zeta_{mn}^{v})_{m,m} = (\mathscr{H}_{mn}^{v})_{m,m}/2\omega_{mn} = [(A_{mn}^{a})_{memb}(S_{mn}^{S})_{memb}\,\mathcal{G}_{v}\Delta_{mnk}]/2\omega_{mn}$

$$= (\frac{h\,^{s}e_{31}}{mn\epsilon_{33}})(-\frac{4e_{31}}{\rho hR\beta*L})\left[\frac{\beta*L}{mn\pi^{2}}\frac{1}{R\pi^{2}}\right]\mathcal{G}_{v}\Delta_{mnk}/2\omega_{mn}\ ,\qquad(12a)$$

ii) $(\zeta_{mn}^{v})_{b,b} = (\mathscr{H}_{mn}^{v})_{b,b}/2\omega_{mn} = [(A_{mn}^{a})_{bend}(S_{mn}^{S})_{bend}\,\mathcal{G}_{v}\Delta_{mnk}]/2\omega_{mn}$

$$= (\frac{h\,^{s}e_{31}}{mn\epsilon_{33}})(\frac{4e_{31}}{\rho hR\beta*L})\,r^{a}r^{s}\left[(\frac{mR\beta*}{nL})+(\frac{nL}{mR\beta*})\right]\left[(\frac{m}{L})^{2}+(\frac{n}{R\beta*})^{2}\right]$$

$$\cdot\,\mathcal{G}_{v}\Delta_{mnk}/2\omega_{mn}\ ,\qquad(12b)$$

iii) $(\zeta_{mn}^{v})_{m,b} = (\mathscr{H}_{mn}^{v})_{m,b}/2\omega_{mn} = [(A_{mn}^{a})_{memb}(S_{mn}^{S})_{bend}\,\mathcal{G}_{v}\Delta_{mnk}]/2\omega_{mn}$

$$= (\frac{h\,^{s}e_{31}}{mn\epsilon_{33}})(\frac{4e_{31}}{\rho hR\beta*L})\,\frac{r^{s}}{R\pi^{2}}\left[(\frac{mR\beta*}{nL})+(\frac{nL}{mR\beta*})\right]\mathcal{G}_{v}\Delta_{mnk}/2\omega_{mn}\ ,\qquad(12c)$$

iv) $(\zeta_{mn}^{v})_{b,m} = (\mathscr{H}_{mn}^{v})_{b,m}/2\omega_{mn} = [(A_{mn}^{a})_{bend}(S_{mn}^{S})_{memb}\,\mathcal{G}_{v}\Delta_{mnk}]/2\omega_{mn}$

$$= (\frac{h\,^{s}e_{31}}{mn\epsilon_{33}})(\frac{4e_{31}}{\rho hR\beta*L})\,\frac{r^{a}}{R\pi^{2}}\left[(\frac{mR\beta*}{nL})+(\frac{nL}{mR\beta*})\right]\mathcal{G}_{v}\Delta_{mnk}/2\omega_{mn}\ .\qquad(12d)$$

If the sensor and the actuator layers are symmetrically laminated on the bottom and top surfaces of the cylindrical shell, the moment arms $r^{s} = -r^{a}$. Thus, the controlled coupling damping ratio components $(\zeta_{mn}^{v})_{m,b}$ and $(\zeta_{mn}^{v})_{b,m}$ are of equal magnitudes and opposite signs and these coupling effects cancel out each other in the damping expression.

PARAMETRIC STUDIES OF SENSITIVITIES AND VIBRATION CONTROL

Based on the derivations and analysis presented previously, a systematic study of sensor sensitivities and actuator effectiveness is presented in this section. Parametric analysis of design parameters, such as sensor/actuator thickness, shell thickness, shell curvatures, shell sizes, is conducted to evaluate sensing/actuation characteristics of segmented shell sensor patches on a simply supported piezoelectric laminated cylindrical shell. Standard dimensions are shell length L = 0.1m, shell curvature angle $\beta^{*} = \pi/2$, shell radius R = 0.05m, elastic lamina thickness $h_{2} = 0.0005$m, and piezoelectric lamina thickness $h_{1} = 20\mu$m. The original modal damping is 1%. Detailed material properties are listed in Table 1. Note that the distributed sensor and actuator layers (the 1st and 5th, respectively) are quarterly segmented into the same patch patterns.

Table 1 Material properties

Properties		Graphite/epoxy	PVDF	Units
Young's modulus	Y_x	181.0	1.6	
	Y_y	10.3	1.6	GPa
	Y_z	10.3	1.6	
Shear modulus	G_{yz}	2.87	0.62	
	G_{xz}	7.17	0.62	GPa
	G_{xy}	7.17	0.62	
Poisson's ratio	μ_{yz}	0.33	0.29	
	μ_{xz}	0.28	0.29	
	μ_{xy}	0.28	0.29	
Density ρ		1600	1800	kg/m^3
Electric permittivity ϵ_{33}			106×10^{-12}	F/m
Piezoelectric const. d_{31}			6.0×10^{-12}	C/N or m/V
e_{31},			9.6×10^{-3}	C/m^2

Case 1: Sensor/Actuator Thickness

In this case, sensor thickness h_1 are evaluated at $h_1 = 10, 20, 30, 40, 50 \mu m$, respectively. Figure 4 shows natural frequencies of these configurations, in which m denotes the half–wave number in the longitudinal direction and n is the half–wave number in the circumferential direction.
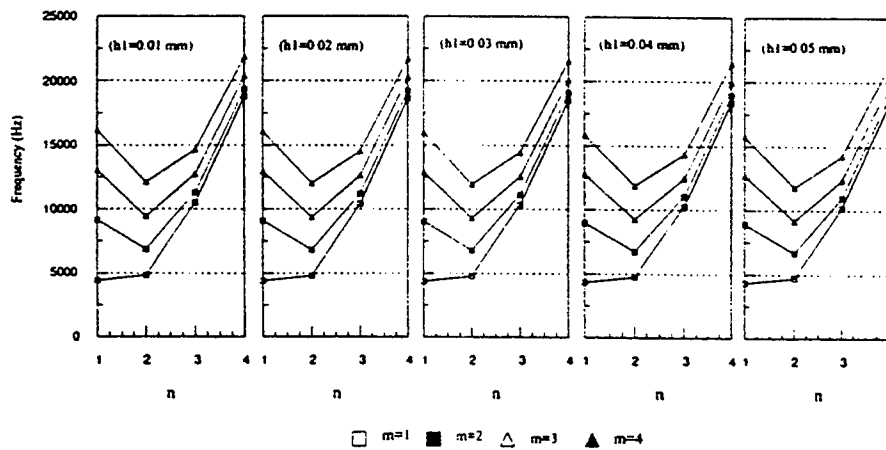
Fig.4 Frequency variations due to thickness changes of sensor layer.

It is observed that natural frequency decreases very little when the sensor thickness h_1 increases because the increment is minimal, i.e., in microns. This small decrease implies that the effect of the increased mass is slightly more significant than that of the increased stiffness of the sensor layer. Modal membrane, bending, and total sensitivities of natural modes m = 1,2,3,4 and n = 1,2,3,4 are plotted in Figure 5. Modal membrane sensitivity linearly increases since the signal is directly proportional to the sensor thickness. Modal bending sensitivity increases due to an increase of h_1 and also the moment arm r^s. Total sensor sensitivity increases with the increase of sensor thickness, especially for lower natural modes. In addition, membrane effects are prominent for lower natural modes and bending effects are prominent for higher natural modes (Tzou, 1993).
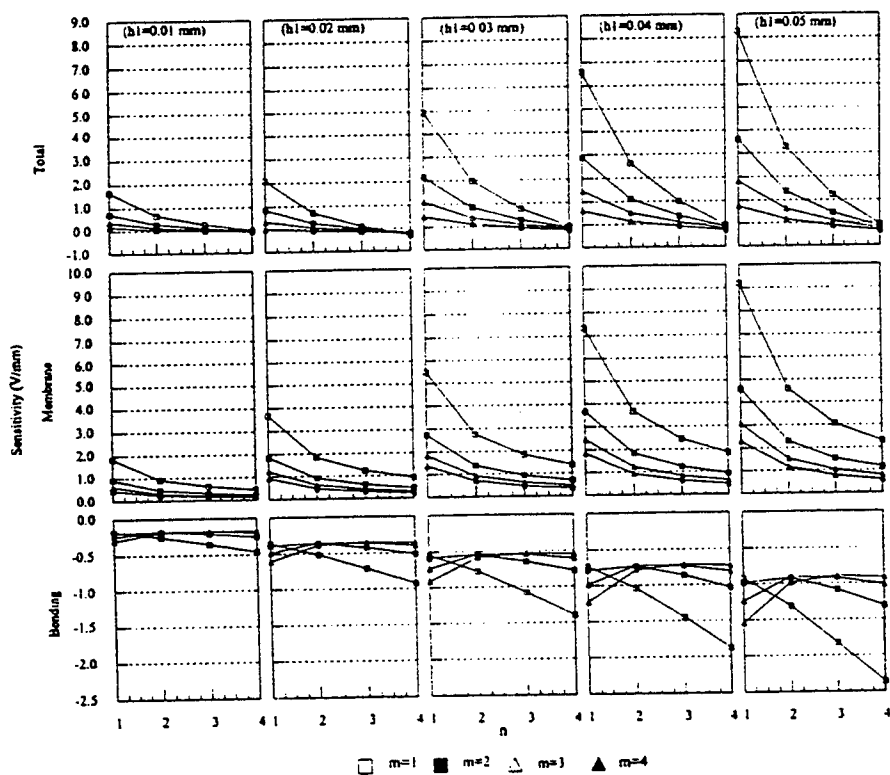
Fig.5 Modal sensitivities of various sensor thickness.

Actuator thickness effects to modal actuation factor, modal velocity feedback factor, and damping ratio in the velocity feedback are evaluated. Analysis results suggest that the major control action comes from the membrane actuation which decreases as the mode number increases. The bending actuation factors remain about the same for all actuator

thickness, which becomes dominating for high natural modes. Since the actuator thickness changes very little and also the actuator material was piezoelectric polyvinylidene fluoride, thickness effect to the total modal actuation factor is relatively insignificant. However, the increased mass does slightly outweighs the increased stiffness and consequently the natural frequency of thicker actuator drops slightly. Controlled damping ratios of the corresponding natural modes are plotted in Figure 6. Since the frequency drops and the modal velocity feedback factor increases as the actuator becomes thicker, the resultant modal damping ratio increases due to enhanced actuation effect. This effect becomes relatively insignificant for higher modes.
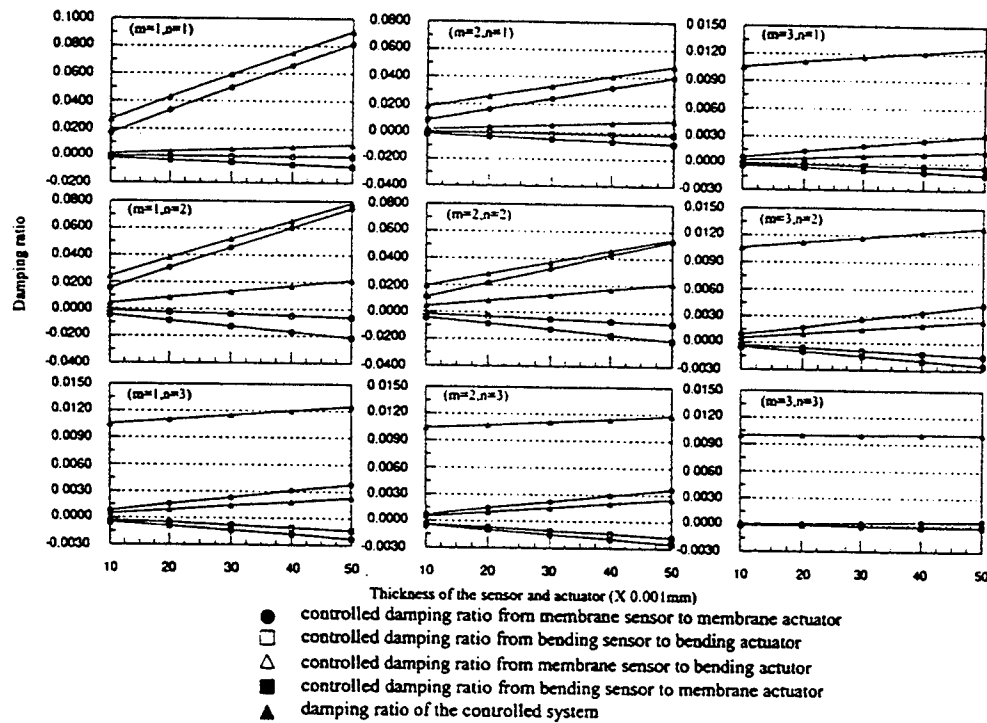
Fig.6 Controlled modal damping ratios of various actuator thickness.

Case 2: Elastic Lamina Thickness

In this case it is assumed that the elastic lamina changes its thickness and $h_2 = 0.2$, 0.3, 0.4, 0.5, 0.6mm, respectively. Natural frequencies and modal sensitivities (membrane, bending, and total) are plotted in Figures 7 and 8.
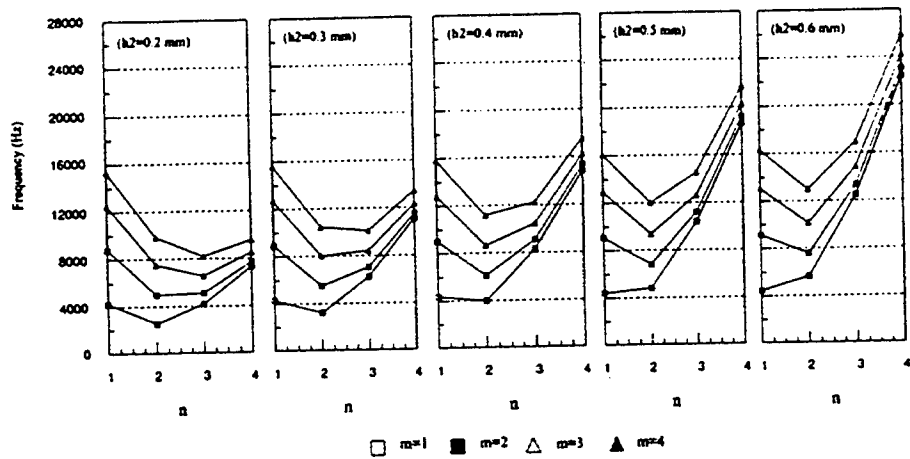
Fig.7 Frequency variations due to thickness changes of elastic laminae.
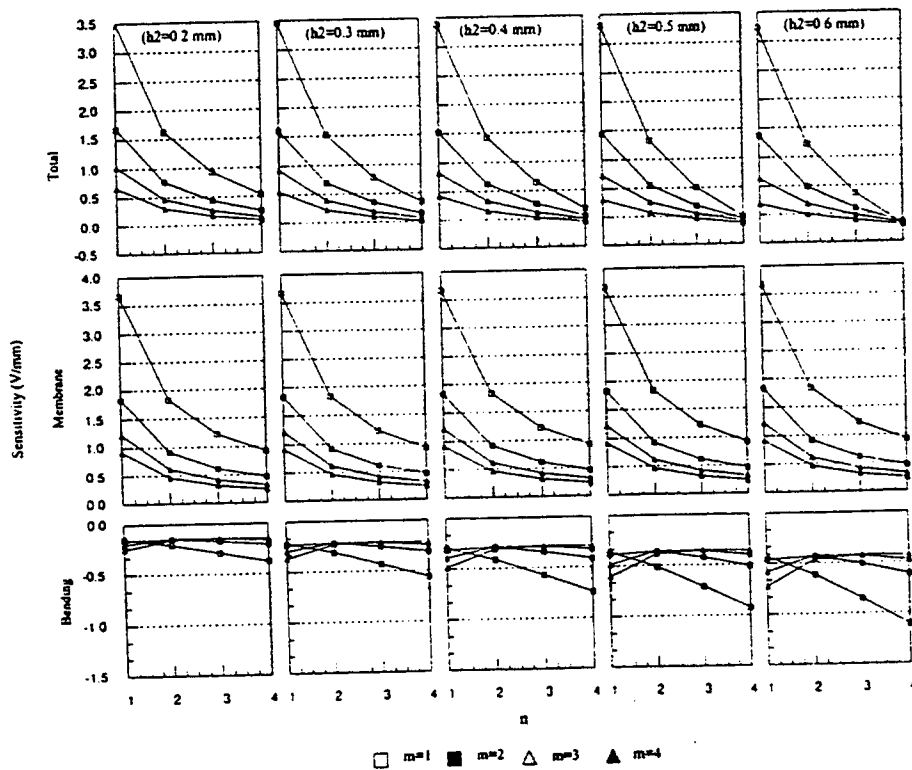
Fig.8 Modal sensitivities of various elastic lamina thickness.

Frequency variations suggest that the increased stiffness due to thickness change contributes to the frequency rise and its effect is more significant than the increased mass.

In addition, as the shell thickness becomes thicker, bending effect becomes dominating in frequency variations. This phenomenon particularly exists in higher natural modes because the kinetic strain energy of higher modes is dominated by bending strain energies while the lower modes are dominated by membrane strain energies. Bending sensitivity increases since the moment arm r^s becomes larger. However, since the thickness change of elastic laminae does not affect its membrane strains, membrane sensitivity remains identical even when the elastic lamina becomes thicker. Total modal sensitivity still decreases, especially for higher modes due to increased bending effects.

Modal damping ratio variations are calculated and plotted in Figure 9. Analytical results suggest that the modal actuation factors and modal damping ratios decrease as the shell thickness increases, due to increased shell bending modulus which is a cubic function of shell thickness. Although the moment control effect increases linearly, the increased shell rigidity outweighs the increased moment control effect, and thus thicker (or stiffer) shells are much more difficult to control.
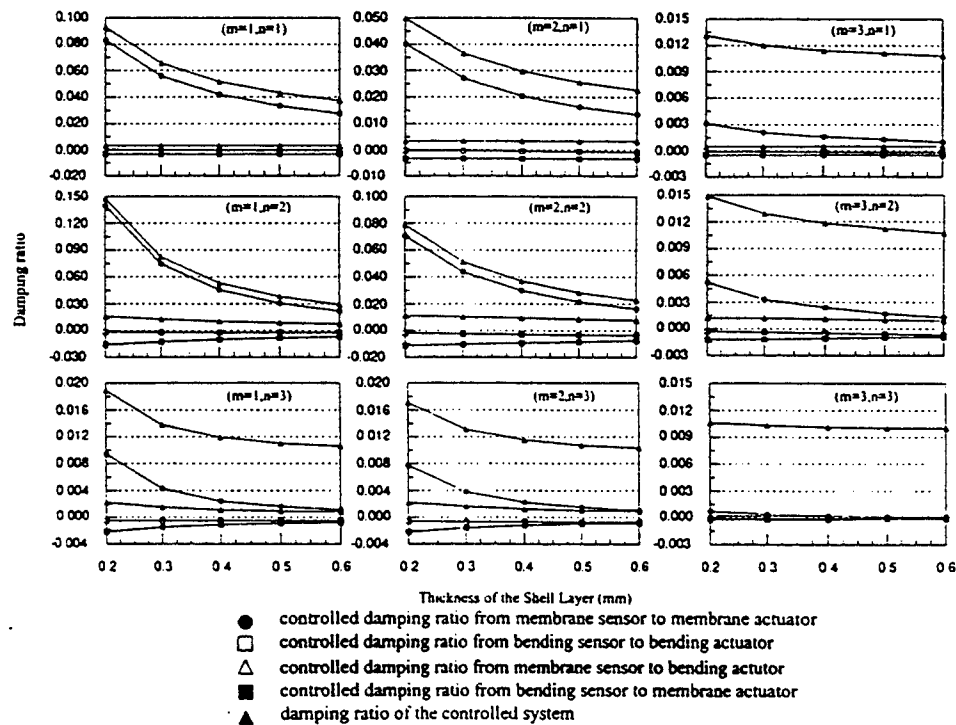
Fig.9 Controlled modal damping ratios of various elastic lamina thickness.

Case 3: Shell Curvature Angles

Adaptive structures with continuous geometry transformations have many potential applications, e.g., optical focusing, flow control, etc. Frequencies and modal sensitivities of a continuous changing cylindrical shell are investigated in this section. It is assumed that the total effective arc (circumferential) length is constant, i.e., $R\beta* = 0.05 \times \frac{\pi}{2}$ m and curvature angles $\beta* = 30°$, $60°$, $90°$, $120°$, $150°$, respectively. Thus, the total shell size remains the same; only its curvature changes. Natural frequencies and modal sensitivities are calculated and plotted in Figures 10. As expected, natural frequencies of shallow shells increase as the mode number increases, and those of deep shells decrease for the first few natural modes and increase as the mode number increases when the curvature becomes significant. This is due to the fact that the kinetic strain energies of lower modes are dominated by membrane strain energies and those of higher modes are dominated by bending strain energies (Tzou and Bao, 1995).
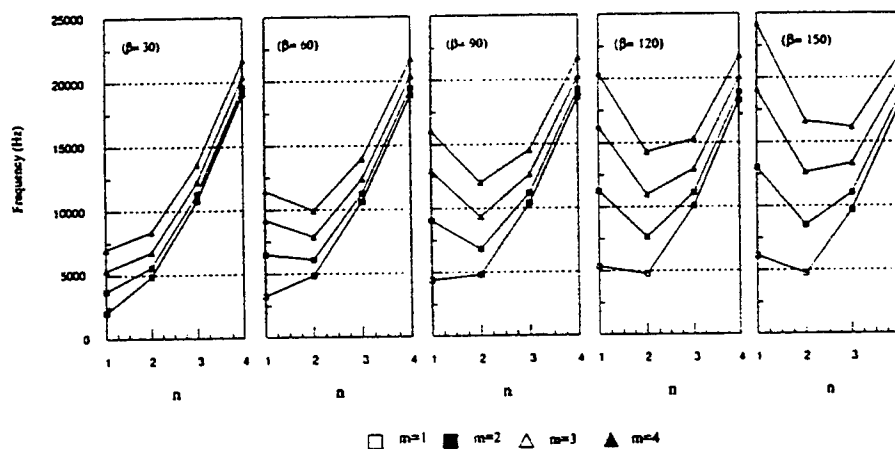
Fig.10 Frequency variations due to curvature changes.

The membrane and transverse motion coupling increases as the shell becomes deeper and deeper. Accordingly, membrane sensitivity increases when the shell curvature increases. However, bending sensitivity remains unchanged since the moment arm never changes over the period of transformation. Overall, the total sensor sensitivity still increases due to the significant increase of membrane sensitivities, Figure 11.
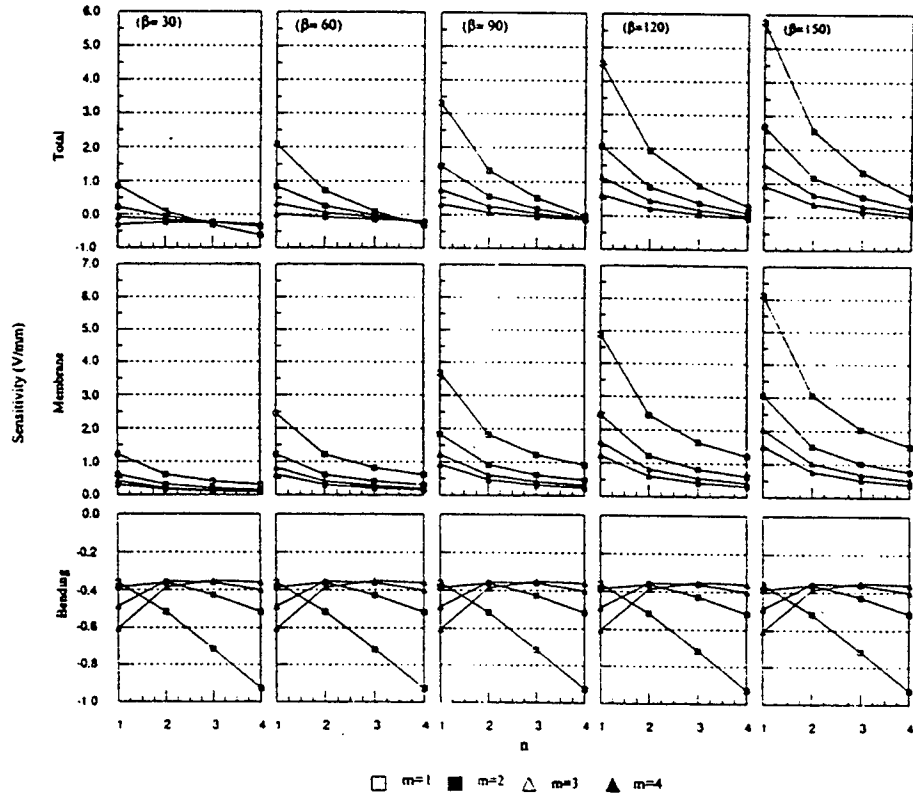
Fig.11 Modal sensitivities of various curvatures.

Continuous curvature transformations are of importance in adaptive (geometry) shells and structures (Tzou and Bao, 1995). Curvature effects to modal actuation factors and modal damping ratios are investigated. Modal actuation factors, modal velocity feedback factors, and modal damping ratios are calculated and evaluated. Analytical results indicate that the membrane actuation effect increases significantly as the shell curvature increases, due to an increased membrane effect in curved shells. Bending control effect remains about the same, since the moment arm remains unchanged. Again, the membrane control effect dominates the lower natural modes as well as the total control effect. Figure 12 summarizes controlled damping ratios. Recall that membrane strain energy dominates in lower natural modes of deep shells and bending strain energy dominates in shallow or zero–curvature shells. For lower natural modes, natural frequency of shallow shells keeps increasing, while natural frequency of highly curved shells drops as mode increases. This frequency variation due to curvature changes also affects the relatively irregular variations of controlled modal damping ratios, i.e., ζ'_{mn}

$$= (\zeta_{mn} + \mathscr{H}^{v}_{mn}/2\omega_{mn}).$$
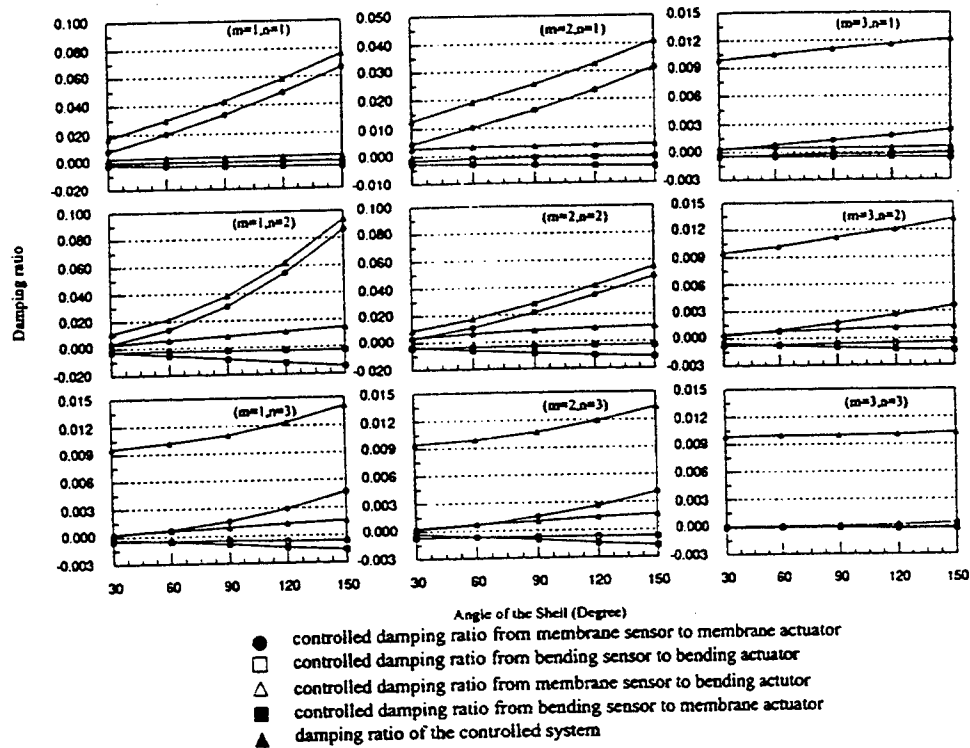
Fig.12 Controlled modal damping ratios of various curvature angles.

Case 4: Shell Sizes

This case is to investigate the sensitivity changes with respect to the size changes. It is assumed that the shell lengths are L = 0.100, 0.125, 0.150, 0.175, 0.200m, and the total arc length $R\beta^* = L$ where $\beta^* = \pi/2$. Since β^* is a constant, shell radius increases when the shell length increases, and accordingly the shell enlarges itself when both length and radius increase. Note that the shell thickness remains unchanged. Again, natural frequencies and modal sensitivities are calculated and plotted. (Due to page limitations, all figures are not presented.) Simulation results suggest that natural frequencies decrease as the shell size enlarges, provided the shell thickness remains unchanged. This is due to an increase in mass and also a decrease in stiffness (i.e., relatively flexible) of the cylindrical shell. Consequently, both membrane strain and bending strain energies become lower, and all modal sensitivities decrease.

Modal actuation factors, modal velocity feedback factors, and controlled damping ratios of these cylindrical shells are calculated and evaluated. Since shell length, radius,

and effective area appear on the denominator of modal actuation factor and controlled damping ratio, increasing both dimensions causes drops in both situations and is insignificant for higher modes.

SUMMARY AND CONCLUSIONS

Spatial characteristics and sensitivities of segmented sensor and actuator patches laminated on a simply supported piezoelectric laminated cylindrical shell were investigated. Detailed sensitivities, actuation factors, feedback factors, and controlled damping ratios were studied with respect to design variables: sensor/actuator thickness, elastic lamina thickness, curvature changes, and shell sizes. Analytical and simulation results of segmented cylindrical sensors and actuators suggested that

1) Thicker sensor layer enhances both membrane and bending sensitivities, and consequently the total sensor sensitivities, especially for lower natural modes. The modal actuation factor slightly decreased as the actuator became thicker. Controlled damping ratio increased due to enhanced actuation effect.

2) Thicker elastic lamina (and shell) possesses stronger bending energies because its increased sectional modulus. Shifts of natural frequencies clearly show the bending energies surpass the membrane energies. Thus, (negative) bending sensitivities increase while membrane sensitivities remain unchanged. The total sensitivities, especially for the higher modes, still decrease due to the increased rigidity or decreased flexibility of the laminated shell. Although the control moment arm is linearly proportional to the shell thickness, the area moment of inertia is a cubic function of the shell thickness. Thus, both modal action factor and modal damping ratio significantly dropped as the shell lamina became thicker.

3) On the other hand, increased shell curvature enhances membrane energies while the bending energies remain identical, as it can be observed in frequency shifts of various curvatures. Consequently, sensor membrane sensitivities significantly increase and the bending sensitivities remain the same. The total sensitivities also increase due to a significant increase in membrane sensitivities. Increasing curvature significantly enhances the membrane effect. Since the dominating control action comes from the membrane control action, curvature increase significantly enhances the modal actuation factor and also the controlled damping ratio.

4) The laminated cylindrical shell becomes relatively flexible as the shell enlarges itself, provided the thickness remains the same. Thus, its natural frequencies drop.

Consequently, both membrane and bending sensitivities fall, and so does the total sensor sensitivity. Modal actuation factor and controlled damping ratio decreased as the shell size enlarged, provided the thickness remained unchanged.

Note that all results were evaluated based on the modal sensitivities and actuations in which gain effects, spatial filtering effects, and temperature effects were not imposed. When spatial filtering is considered, all quadruples of shell natural modes in the quarterly segmented sensor/actuator configuration are eliminated, i.e., they are neither observable nor controllable. In addition, all results were evaluated based on constant piezoelectric coefficients without hysteresis or temperature effects. A detailed and complete report of this study is presented in a review chapter (Tzou, Bao, and Venkayya, 1996).

ACKNOWLEDGEMENT

REFERENCES

Colins, S.A., Miller, D.W., von Flotow, A.H., 1994, "Distributed Sensors as Spatial Filters for Active Structural Control," *J. of Sound & Vibration*, Vol.173.4, pp.471–501.

Detwiler, D.T., Shen, M.–H., and Venkayya, V.B., 1994, "Two–dimensional Finite Element Analysis of Laminated Composite Plates Containing Distributed Piezoelectric Actuators and Sensors," *Proc. 35th AIAA/ASME Adaptive Structures Forum*, pp.451–460, Hilton Head, SC, April 21–22, 1994.

Gu, Y, Clark, R.L., Fuller, C.R., and Zander, A.C., 1994, "Experiments on Active Control of Plate Vibration Using Piezoelectric Actuators and Polyvinylidene Fluoride Modal Sensors, ASME *Journal of Vibration & Acoustics*, Vol.116, pp.303–308.

Hubbard, J.E. and Burke, S.E., 1992, "Distributed Transducer Design for Intelligent Structural Components," *Intelligent Structural Systems*, Tzou, H.S. and Anderson, G.L. (Ed.), Kluwer Academic Publishers, Dordrecht/Boston/London, pp.305–324.

Lee, C.K., 1992, "Piezoelectric Laminates: Theory and Experimentation for Distributed Sensors and Actuators," *Intelligent Structural Systems*, Tzou, H.S. and Anderson, G.L. (Ed.), Kluwer Academic Pub., Dordrecht/Boston/London, pp.75–167.

Qiu, J. and Tani, J., 1994, "Vibration Control of a Cylindrical Shell Using

Distributed Piezoelectric Sensors and Actuators," *Proceedings of the Second International Symposium on Intelligent Materials*, pp.1003–1014.

Suleman, A. and Venkayya, V.B., 1994, "Flutter Control of an Adaptive Composite Panel," *Proc. 35th AIAA/ASME Adaptive Structures Forum*, pp.118–126, Hilton Head, SC, April 21–22, 1994.

Sumali H. and Cudney, H., 1993, "Segmented Two–dimensional Modal–filtering Sensors," DE–Vol.61, *Vibration and Control of Mechanical Systems*, pp.59–66. Design Technical Conference, Albuquerque, NM, September 1993.

Tzou, H.S., 1993, *Piezoelectric Shells (Distributed Sensing and Control of Continua)*, Kluwer Academic Pub., Dordrecht/Boston/London.

Tzou, H.S. and Anderson, G.L. (Editors), 1992, *Intelligent Structural Systems*, (ISBN No.0–7923–1920–6), Kluwer Academic Publishers, Dordrecht/Boston, August 1992.

Tzou, H.S. and Bao, Y., 1995, "Dynamics and Control of Adaptive Shells with Curvature Transformations," *Shock and Vibration Journal*, Vol.2, No.2, pp.143–154.

Tzou, H.S., Bao, Y., and Venkayya, V.B., 1996, "Micro Electromechanics and Functionality of Segmented Cylindrical Shell Transducers," *Structronic Systems: Smart Structures, Devices and Systems*," Editors: Tzou, et al., World Science Pub. NJ.

Tzou, H.S. and Fu, H.Q., 1994, "A Study of Segmentation of Distributed Piezoelectric Sensors and Actuators, Parts 1 and 2," *Journal of Sound & Vibration*, Vol.172, No.2, pp.247–276.

Tzou, H.S. and Fukuda T. (Editors), 1992, *Precision Sensors, Actuators, and Systems*, Kluwer Academic Pub., Dordrecht/Boston/London, December 1992.

Tzou, H.S. and Hollkamp, J.J., 1994, "Collocated Independent Modal Control with Self–sensing Orthogonal Piezoelectric Actuators (Theory and Experiments)," *Journal of Smart Materials and Structures*, Vol.3, pp.277–284.

Tzou, H.S. and Tseng, C.I., 1991, "Distributed Modal Identification and Vibration Control of Continua: Piezoelectric Finite Element Formulation and Analysis," ASME Journal of Dynamic Systems, Measurements, and Control, Vol.113, No.3, 500–505.

Tzou, H.S., Zhong, J.P., and Natori, M.C., 1993, "Sensor Mechanics of Distributed Shell Convolving Sensors Applied to Flexible Rings," *ASME Journal of Vibration & Acoustics*, Vol.115, No.1, pp.40–46.

Tzou, H.S., Zhong, J.P., and Hollkamp, J.J., 1994, "Spatially Distributed Orthogonal Piezoelectric Shell Actuators (Theory and Applications)," *Journal of Sound & Vibration*, Vol.177, No.3, pp.363–378. (Chap–SegShl.Dst195.Intg395.RDL–Rpt/WpFi295)

EVALUATION OF:
OPEN ARCHITECTURE MACHINE TOOL CONTROLLERS
&
AGILE MANUFACTURING

William J. Wolfe
Associate Professor
Computer Science and Engineering

University of Colorado at Denver
Campus Box 109
Denver, Colorado 80201
wwolfe@cse.cudenver.edu
(303) 556 - 2358

Final Report for:
Summer Faculty Research Program
Wright Laboratory
Manufacturing Technology Directorate

Sponsored by:

July, 1995

EVALUATION OF:
OPEN ARCHITECTURE MACHINE TOOL CONTROLLERS
&
AGILE MANUFACTURING

William J. Wolfe

Associate Professor

Computer Science and Engineering

University of Colorado at Denver

Abstract

This report provides an evaluation of the Open Architecture Machine Tool Controllers and Agile Manufacturing programs. The most interesting result is that the concept of the *open architecture* with interoperable/portable *agents*, developed for the Open Architecture Program, can also provide a much needed structure for the Agile Manufacturing Program. Both programs deal with a wide variety of functional components that must *evolve* in an uncertain technological environment. Specifying the design of these components is greatly complicated by a rapidly changing technology base. It is almost impossible to predict the performance of a particular module without testing it over a period of time in a wide range of operational environments (*portability* and *interoperability*). The *open architecture* strategy provides the necessary interface information so that many independent developers can evaluate their modules. The strategy does not commit to any particular design, it just provides the *playing field* (i.e.: interface specifications). The clients submit their modules for testing: performance? interoperability? portability? The *best* modules will survive the test of time. This strategy encourages many *third party players*, thus tapping a diverse network of developers. Furthermore, the rapidly emerging client/server networking technology (internet, world-wide web, etc.) will provide the necessary communications infrastructure for the many geographically dispersed developers. The Open Architecture Machine Tool Controllers program has come a long way in identifying and implementing these concepts. The Agile Manufacturing program can benefit greatly by following this lead.

Table of Contents

1. Introduction

The Manufacturing Technology Directorate (WL/MT) asked me to provide an evaluation of the Open Architecture Machine Tool Controllers (OA) and the Agile Manufacturing (AM) programs as part of my summer faculty position at Wright Laboratory, sponsored by AFOSR. My background consists of more than 20 years of AI and Robotics projects, industry and university, and I am currently a Professor of Computer Science at the University of Colorado at Denver, where I teach graduate courses in Artificial Intelligence, Neural Networks, Algorithms for Planning/Scheduling, and Expert Systems. I was encouraged to: "tell us what you think".

2. Summary

In this report the Open Architecture and Agile Manufacturing programs are independently evaluated, and then comparisons are made at the end. More effort was put into the OA program since it was the first task and it had the most documentation. The most important recommendation in this report is that the AM program should *piggyback* on the achievements of the OA program: an *open architecture* with independently developed *agents*, supported by a client/server networking infrastructure. This is the best strategy for both the OA and the AM programs for the following reasons:

- rapidly changing technology;
- diverse and dispersed third party developers;
- emerging communications network infrastructure.

A rapidly changing technology base makes it difficult, if not impossible, to quickly evaluate a particular method, architecture, device, software module, or any other design concept. Advances in computer technology, factory machines, sensors, networks, and databases have the potential to change the directions of the technology in a heartbeat. This makes it difficult for developers to predict how well their module will work (portable? interoperable?, etc.). To further complicate matters, there are potentially *thousands* of independent developers. We need to support these 3rd party developers, because they are a highly adaptive, reliable, and affordable source of technology. The open architecture strategy does just that. It offers interface specifications and manages the evolution of suggested modules (configuration control), while providing the necessary testing ground. The best modules will survive the test of time if they show high degrees of performance, portability, and interoperability. Finally, the emerging client/server networking technology

(internet, world-wide web, etc.) will directly facilitate the interactions between the many geographically dispersed users and developers.

With this said, it is still no easy matter to define a specific open architecture for the Agile Manufacturing Program, but the overall strategy is clear. From the top-down we want to encourage portable, interoperable, components and avoid cumbersome, fragile, modules[1]. We want to encourage as many third party player as possible and provide a testing ground for new modules, while supporting the networking infrastructure. The Open Architecture Program is forging ahead in this direction, so lessons learned can be passed along to the AM program.

3. Open Architecture Machine Tool Controllers Program (OA)

With regard to the Open Architecture Program (OA) program, I quickly discovered that Lonnie Burnett (Lawrence Associates Inc.) is the resident expert, covering all aspects of the program. It is clear that both the Open Architecture and the Agile Manufacturing programs are populated with some of the best people in the industry, both managerial and technical, and represent unprecedented levels of industry and university teamwork. The OA program is breaking new ground in the controller industry and, because of the innovative *open architecture* approach, with the help of the *agent* concept, is having wide ranging side effects, especially in areas where large systems of *interoperable* software components are required[2].

Most Important OA Issues:

- Controller R&D vs. creation of a "new business";

  (Research vs. Business)

- The balance of *secrecy* and *openness*;

  (Shared vs. Proprietary)

- Delay in defining the initial straw-man Open Architecture;

  (Top Down vs. Bottom-Up)

Clearly, establishing a new business is very different from participating in controller R&D. The OA program has elements of both. This creates a significant management challenge

---

[1] Object-Oriented developments mirror these same goals.

and encourages a continuing confusion over the goals of the program. Combine that with ill-defined boundaries between *open* and *secret* contractor designs and you have a management time bomb ticking away. It is relatively simple to say "vendors will openly define the *interfaces* for their components (agents), but *internal* designs are proprietary". But, the meaning of "interface" is not crystal clear, and is especially vague in the absence of an "agreed upon Open Architecture". The delay in defining the initial open architecture is probably traceable to the ambiguous boundaries between *open* and *secret* which, in turn, is traceable to vague program goals.

Despite these reservations, I am encouraged by Lonnie's comment:

> *"Careful monitoring will be needed to assure that the latent competitiveness of the team members does not destroy team cooperation".*

Although the program is faced with significant challenges, the program is also ensured a high measure of success because it has already identified the critical technological path (open architectures and reconfigurable agents). Success will be even greater when the program capitalizes on the rapidly emerging communications networking (internet, www, etc.). The final success of the program, however, will be determined by the degree of cooperation between contractors.

The OA program will put the USA at the top of the controller industry, and will have far reaching effects on future developments of large software systems.

Finer Points About the OA Program:

1. The program would benefit from a concise, agreed upon, statement of the goals.
2. Sharing of information between competitors is critical and must be stimulated.
3. Encouraging *openness* while respecting *secrecy* is problematic.
4. Many potential 3rd party vendors should be brought in as soon as possible.
5. Focus on *demonstrations* of interoperability *between* sites.
6. A formal means of communicating "lessons learned" would be helpful.
7. OASYS should probably delegate more authority to the integrators.
8. Potential Site dissatisfaction is a significant liability.

---

[2] Such system are, in fact, representative of the wave of the future.

9. The meaning of "Upgrades" should be nailed down.

10. The program is struggling to define an initial straw-man open architecture.

11. "Openness" is an *ideal* goal, not expected to be a reality.

12. The term "agent" is becoming a catch-all.

13. The Title III "new business" should be clarified.

Lonnie Burnett stated the goal of the program as:

"The goal of the program is to develop an open architecture supported by the introduction of controllers and add-ons as products from multiple vendors."

But there is also a <u>business</u> goal. The OASYS Architecture Specifications document (July 6, 1995), states the goal the following way: *"The goal of the OASYS project is to change the market dynamics such that the user rather than the controller provider determines the make-up of the controller - with an OASYS controller, the controller provider gets a vote not a veto on how the control is composed."* The SOW puts a different spin on the goals: *"The overall objective of this project is to establish a cornerstone for viable, long-term world-class domestic manufacturing capabilities for open architecture machine tool controllers."* The SOW goes on to provide many variations on the goals of the project, ranging from business to technical. The main point: it would be nice to agree on a *primary* goal of the program (maybe one technical and one business goal).

The tension between "new business" and "R&D" is accentuated by the "upgrades". The open architecture should facilitate the upgrades, and in a way, the upgrades *demonstrate* the benefits of openness. This will undoubtedly lead to better vendors (more knowledgeable, more customers, etc.) but how this supports a separate, distinct, new business escapes me.

Although the concepts of "openness" and "agent" provide a much needed layer of abstraction, they can become barriers to progress. Much of the ideal-ness inherent in these concepts will not survive the reality of implementation. Consequently, it will be difficult for engineers to maintain a crisp design-to-reality documentation trail. For these reasons it seems necessary to accept some vagueness in the initial Open Architecture design. Where you draw line, and what constitutes an acceptable *initial* Open Architecture design, is best left to the likes of Lonnie Burnett, but I would encourage an acceleration of the process. The program is progressing in a "top down" manner at the moment: emphasis on up-front

design. It is now desirable to shift to a "bottom up" perspective: implementation, lessons learned, and refinement.

It would be wise to apply a more hierarchical management structure because OASYS could easily be overwhelmed with site-specific development details. Figure 1 provides a simple hierarchy that would shield OASYS from some of these specifics, without endangering the program.

Figure 1: A more hierarchical structure is recommended.
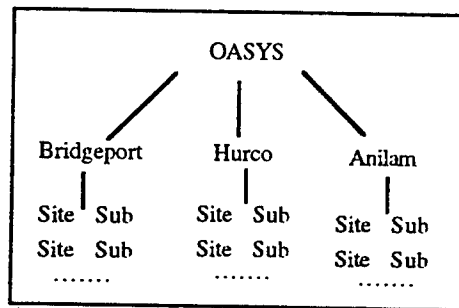
The strength of the OA program is in the identification and development of the open architecture, agent-based, design. I would add to that the power of the world-wide web. The basic structure that emerges it shown in Figure 2.
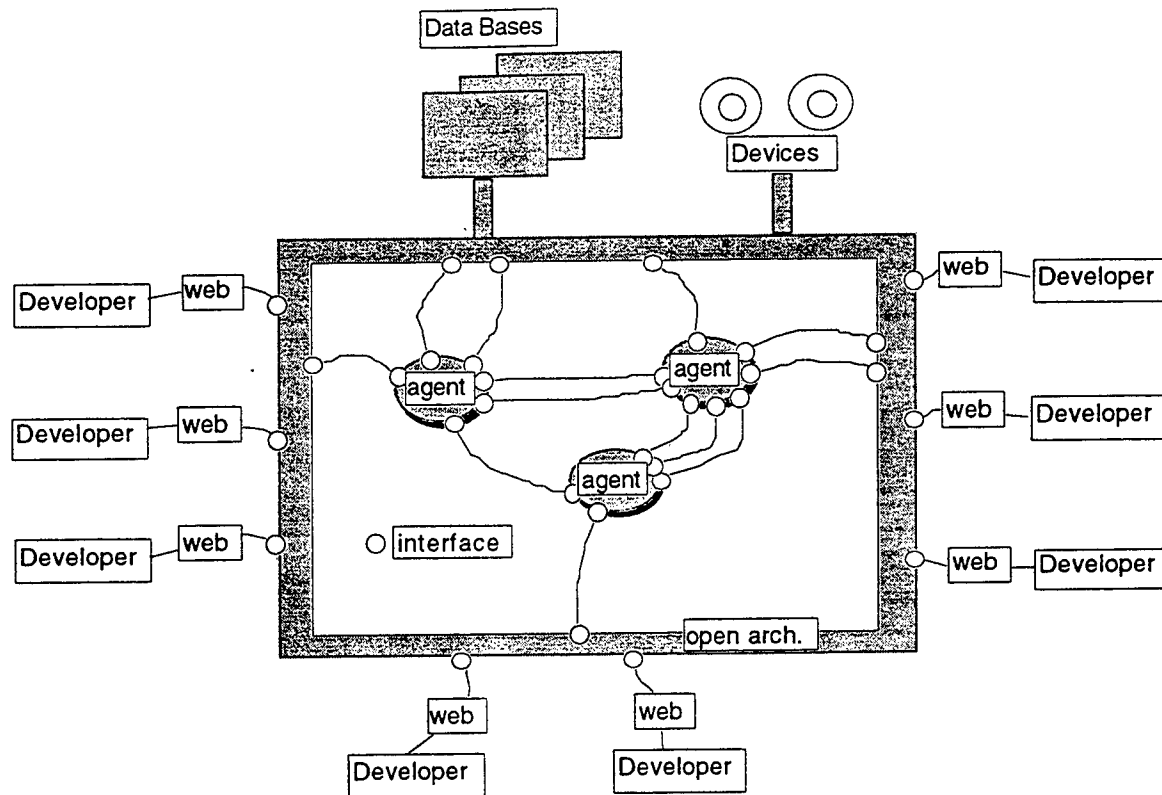
Figure 2: *The basic structure that emerges from the integration of the open architecture, agent-based, and www client/server concepts.*

4. Agile Manufacturing

The Agile Manufacturing Program is involved in many aspects of advanced factory concepts. Here I will be brief and emphasize two technical articles that I reviewed:

[1] *Toward the Development of Flexible Mixed-Initiative Scheduling Tools*, S. Smith and O. Lassila, Proceedings of the 1994 ARPA Planning Workshop, Tucson, AR, February, 1994.

[2] *Development of an Integrated Process Planning/Production Scheduling Shell for Agile Manufacturing*, N. Sadeh, T. Laliberty, R. Bryant, S. Smith, Proceedings of IJCAI-95, Workshop on Intelligent Manufacturing.

These articles provide an excellent overview of the problems that realistic scheduling systems must deal with, such as:

- Vague criteria;

- Rapidly changing conditions (uncertainty, unexpected events, policy changes);

- Combinatorial explosion.

With these difficulties in mind the authors recommend an incremental, iterative, or reactive approach to scheduling. Reference [1] expresses it this way:

"Schedule construction in practice tends to be a dynamic reactive process. An initial schedule is built, problematic or unsatisfactory aspects are identified, requirements relaxed or strengthened ..., schedule modifications are made and so on."

The *plan* is continually evaluated, looking for bottlenecks and particular constraint problems, and *fixes* are made using specialized heuristics and relaxation strategies. The plan is also updated to account for new information and changing conditions. [1] points out that small or local changes to the plan are preferable at each iteration, so as to maintain a comfortable degree of *continuity* and *stability* in a highly dynamic environment. Figure 3 provides a sketch of these concepts.
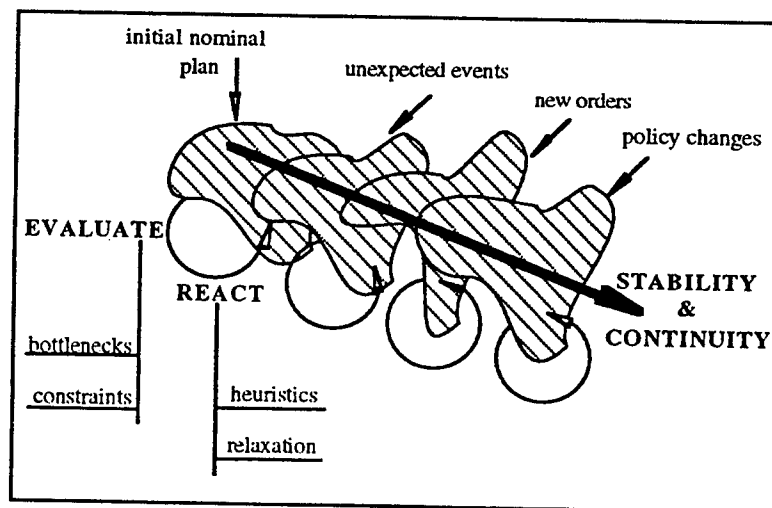
Figure 3: *The basic strategy is to react to a series of disturbances, while considering many "what if" scenarios, and trying to maintain some continuity in the plan as it is incrementally improved.*

There are two other concepts that I would toss into these descriptions: *progressive refinement* and *priority aging*.

Progressive Refinement: Progressive refinement begins with a rough *out day* schedule, and progressively refines it until it is a valid schedule at *day 0* (i.e.: operations) This opens up a variety of hierarchical possibilities. The further the out day the more crude the "schedule" can be, that is, all the hard constraints do not have to be satisfied; for example, there may be some *overbooking*, an unspecified machine or tool, etc. As the *near day* approaches, details can be added and adjustments made, while simultaneously accounting for unexpected events (canceled orders, priority changes, equipment failure, etc.). The schedule may be modified right up to the time of operation. This is also typical of how business trips, and vacations are planned. The main theme is to avoid wasting efforts on details that will be preempted by subsequent events (a type of *least commitment* strategy).

Figure 4: Progressive refinement adds the details as the day of operation approaches, like landing a plane.
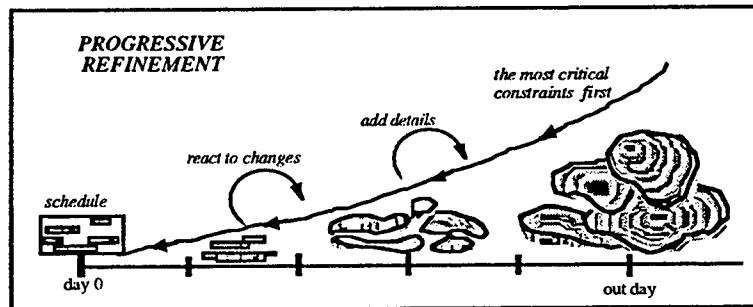
- *Priority Aging* : when a task is scheduled a *commitment* is made that can spawn many contingent plans, possibly reaching out to other organizations beyond the domain of the current schedule. If the task is suddenly bumped, or moved, the effect can propagate in unexpected and undesirable ways (*ripple*). Intuitively, a scheduled task *grows roots* as it sits on a schedule waiting for execution. It is therefore wise to include a factor that raises the *effective priority* of an incumbent task the longer it sits on a schedule, making it more and more difficult to move or bump.

I would also reinforce the following concepts that were brought up in one way or another in [1] and [2]:

- *Stability*: this concept refers to the degree of task movement we are willing to tolerate to incorporate a new task, or to account for a machine failure, etc. It may be possible to get a new task on the schedule without *bumping* (i.e.: *off* the schedule) any task, but several tasks may have to be modified. It is not wise to be continually shuffling several tasks around (i.e.: avoid *nervous* scheduling) since, for example, a subtle system safety factor

might be overlooked, or the humans who interact with the system might become thoroughly confused. Therefore a limit should be put on the amount of disruption introduced at any one time (i.e.: let the *ripples* settle).

• *Opportunistic Planning*: unexpected changes usually have an adverse effect on the quality of a schedule, but in some situations a change creates previously unconsidered opportunities. Be on the look out for such *free* opportunities, and capitalize on them.

What emerges from this analysis is an appreciation for the depth of the difficulties, and the need for *decision support* tools, relying on the user to provide the majority of the intelligence. It is wise that both [1] and [2] anchor their research in specific domains: Transportation Scheduling [1], and Raytheon Factory Scheduling [2]. However, defining the level of user/system interaction is not simple. For example, bridging the gap between number-crunching algorithms, whose details should be hidden from the user, and discrete decision points, where the user should be presented with clear choices, is a *major* research area.

[1] attacks this major research area by breaking the problem into two pieces (I quote):

- a *decision making* component, responsible for making choices among alternative scheduling decisions and retracting those that have since proved undesirable, and

- a *constraint management* component, whose role is to propagate the consequences of decisions and incrementally maintain a representation of the current set of feasible solutions (detecting inconsistent solution states when they arise).

The concept is further refined by splitting the *decision making* step into two steps:

- *action formulation:* concerned with isolating a particular subproblem and

- *action execution:* solutions to these subproblems.

The author ([1]) provides a few examples from the transportation scheduling domain to help clarify the meanings of these terms. Reference [2] acknowledges many of the same issues as [1] (mix-initiative, decision support, etc.) but takes a more generalized view of

the dynamics of the emerging architecture, and describes a *blackboard* approach to the integration of diverse knowledge sources. The theme is a "common representation for exchanging process planning and production scheduling information", while "enabling the user to interactively explore a number of tradeoffs". The resulting system[3] will be demonstrated at the Raytheon Andover manufacturing facility. Again, it seems wise that the research is anchored in a demonstration facility.

It is pointed out in [2] that there are various "islands of integration" emerging in the CAD/CAM world. These *islands* reflect the need to tighten up the connections between customer requests, marketing evaluations, process planning, and production scheduling. With these connections tightened, a company can rapidly respond to new products, new suppliers, and new demand levels. The author goes on to say: "... building the bridges between these islands is without doubt the next major hurdle in developing Computer Aided Manufacturing environments capable of effectively supporting Agile Manufacturing practices." This is an interesting way of seeing the big picture. It sees the technology as emerging over a period of time as linkages are refined, common information structures are identified, modules are developed by diverse groups, tested, etc. The important point here, I think, is that no one can predict the exact direction of these developments, and the systems (islands) will *emerge* as they are proven to be useful.

This is the same picture that surfaced from about 3 years of intensive research into the Open Architecture program, originally called Next Generation Controller. The result was the *open architecture* with independently developed reconfigurable *agents*. In this sense, the blackboard structure defined in [2] is a small scale version of the big picture, and creates a nice local/global duality in system design. The power of this overall approach is amplified by several orders of magnitude when the potential of advanced communications networking (internet, www, etc.) are brought into the picture.

5. Conclusions

This report highlights the similarity between the Open Architecture and the Agile Manufacturing Programs, pointing out that the OA program is a step ahead of the AM program, and that OA should follow the basic principles of the open architecture, agent based design strategy, while adding on an intense effort to use advanced communications such as the internet and world-wide web. Along the way there will be many false starts and a resistance to change but things will smooth out with a steady progress because of the

---

[3] Integrated Process Planning/Production System (IP3S).

many developers who are willing and able to provide affordable components. The critical need however, is for the development of a common infrastructure. This puts the burden on the configuration control aspect of the problem. Who, or what agency, is going to step forward to act as the "manager" of these developments? This involves the establishment of interface specifications and the promulgation of changes/updates. Additionally, there is a need for development of measures of portability, reconfigurability, interoperability, etc., so that the better modules will be recognized.

Various managerial and operational personnel, faced with a wide variety of day-to-day planning and scheduling problems, will be the *users*. They will evaluate the various tools, made available through the network, as the open system evolves. They should have access to a library of tools, and as time goes on, they become the best judges. If a certain tool is only useful for a very specific problem that rarely occurs, then the tool will tend to die on the vine (with proper configuration control), and a tool that runs poorly but is for some reason used quite frequently then the demand for an improvement will become known in the developer community and new versions of the tool will quickly emerge. This, of course, is how hardware and software is currently being developed in the PC market.

6. Acknowledgments

7. References

[1] *Toward the Development of Flexible Mixed-Initiative Scheduling Tools*, S. Smith and O. Lassila, Proceedings of the 1994 ARPA Planning Workshop, Tucson, AR, February, 1994.

[2] *Development of an Integrated Process Planning/Production Scheduling Shell for Agile Manufacturing*, N. Sadeh, T. Laliberty, R. Bryant, S. Smith, Proceedings of IJCAI-95, Workshop on Intelligent Manufacturing.

[3] *OASYS Architecture Specification for the Open Architecture Machine Tool Controllers*, OASYS Group, Inc., contract F33733-95-C-1018, July 6, 1995.

[4] *Next Generation Controller Specification for an Open Systems Architecture Standard*, Martin Marietta Astronautics, Denver, CO, Sept. 28, 1994.

[5] *Virtual Manufacturing Technical Workshop 25-26 October 1994, Technical Report*, Lawrence Associates Inc., Jan. 26, 1995.

A FINITE-VOLUME, TIME-DOMAIN FORMULATION
FOR WIDE-BAND RCS PREDICTION

Jeffrey L. Young
Assistant Professor
Department of Electrical Engineering

University of Idaho
Moscow, ID 83844-1023

A FINITE-VOLUME, TIME-DOMAIN FORMULATION FOR WIDE-BAND RCS PREDICTION

Jeffrey L. Young

Assistant Professor

Department of Electrical Engineering

University of Idaho

Abstract

A finite-volume, time-domain code for radar cross-section (RCS) prediction (authored by J. S. Shang) was further developed. Specifically, the modified code is presently capable of simulating wide-band input pulses and calculating RCS information over a band of frequencies. The post-processing RCS routine was completely rewritten; the new routine is based upon the equivalence theorem for far-field power observations. The scattering object under investigation was a perfectly conducting sphere; various electrical sizes were considered. The data obtained from the code was in agreement with the exact Mie solution.

In addition to code development, a new numerical boundary condition for collocated schemes was postulated. This new condition shows explicitly the relationships between the surface charge, surface currents and the tangential electric field. By having such a condition, we circumvent the extrapolation process used currently. Validation of this concept is in progress.

A FINITE-VOLUME, TIME-DOMAIN FORMULATION
FOR WIDE-BAND RCS PREDICTION

Jeffrey L. Young

Assistant Professor

Department of Electrical Engineering

University of Idaho

Introduction

The prediction of an object's radar cross section (RCS) via a numerical code has received considerable amount of attention in the past two decades by both the military and civilian communities [1]–[7]. Due to the computer revolution, practical geometries spanning hundreds of wavelengths can now be modeled with accuracy and speed. For example, for scattering bodies spanning up to twenty wavelengths, integral equation and partial differential equation (PDE) solvers can be employed; obversely, physical and geometrical optics methods have proven their worth for optically large objects.

Although each of the aforementioned solvers have their specific advantages as well as disadvantages, this report is concerned with time–domain PDE solvers of the finite–volume type as applied to electromagnetic scattering processes – namely, the determination of an object's RCS. This area of research was the focus of work for the author during his stay at the Wright Patterson Air Force Base, Flight Dynamics Directorate for the summer of 1995.

Several achievements and contributions were made this last summer. In the ensuing section, these achievements are listed explicitly. Subsequent sections provide the specific details.

Achievements and Contributions

A finite–volume, time–domain (FVTD) code, authored by Dr. J. S. Shang, was further

developed and refined. Below is a brief summary of the specific achievements and contributions made to this code; other contributions are also listed.

- A scattered–field formulation for lossy, dielectrically coated bodies was installed into the FVTD code.

- The FVTD code was modified for pulse excitation. Post-processing Fourier routines and source routines were developed.

- A post–processing RCS routine was written; the formulation uses only equivalent electric and magnetic surface currents.

- RCS results were generated and verified for the perfectly conducting sphere; the excitation was a gaussian–pulsed plane wave.

- RCS results were generated for the coated, perfectly conducting sphere; verification of data is still in progress.

- A new theoretical boundary condition for perfect conductors was postulated; this portion of the research is the topic of the proposal "Development of a new numerical boundary condition for perfect conductors," which has been submitted to the Research Defense Laboratories (RDL).

- A paper to the AIAA, New Orleans conference is being written. Dr. Y. Weber of the Flight Dynamics Directorate is the lead author.

Scattered–Field, Finite–Volume Formulation

The characteristic–based, finite volume technique is founded on the principle that Maxwell's equations can be couched in the following strong conservative form [5]–[7]:

$$\frac{\partial \hat{U}}{\partial t} + \frac{\partial \hat{F}}{\partial \xi} + \frac{\partial \hat{G}}{\partial \eta} + \frac{\partial \hat{H}}{\partial \zeta} = -\hat{J} \qquad (1)$$

where $\hat{U}$ is the unknown vector and $\hat{F}, \hat{G}$, and $\hat{H}$ are the fluxes in the ξ, η, ζ directions, respectively. Under the assumption that an appropriate body conformal transformation of the type $\xi = \xi(x, y, z), \eta = \eta(x, y, z)$ and $\zeta = \zeta(x, y, z)$ can be defined with a corresponding Jacobian V, the fluxes, $\hat{U}$ and $\hat{J}$ take on the following meaning:

$$\hat{U} = UV, \tag{2}$$

$$\hat{J} = JV, \tag{3}$$

$$\hat{F} = (\xi_x F + \xi_y G + \xi_z H)V, \tag{4}$$

$$\hat{G} = (\eta_x F + \eta_y G + \eta_z H)V, \tag{5}$$

and

$$\hat{H} = (\zeta_x F + \zeta_y G + \zeta_z H)V, \tag{6}$$

provided that

$$U = [B_x^s, B_y^s, B_z^s, D_x^s, D_y^s, D_z^s]^t, \tag{7}$$

which are scattered field quantities. Since the fluxes are homogeneous functions of order one, it is a simple process to show that $F = AU$, $G = BU$ and $C = HU$, where, for example,

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -\frac{1}{\epsilon} \\ 0 & 0 & 0 & 0 & \frac{1}{\epsilon} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\mu} & 0 & 0 & 0 \\ 0 & -\frac{1}{\mu} & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{8}$$

Here the permittivity $\epsilon = \epsilon(x, y, z)$ and the permeability $\mu = \mu(x, y, z)$ describe the inhomogeneous coating; exterior to the coating they take on their free–space values of ϵ_0 of μ_0, respectively. Given A, B and C, it follows that

$$F = [0, -D_z^s/\epsilon, D_y^s/\epsilon, 0, B_z^s/\mu, -B_y^s/\mu]^t \tag{9}$$

$$G = [D_z^s/\epsilon, 0, -D_x^s/\epsilon, -B_z^s/\mu, 0, B_x^s/\mu]^t \tag{10}$$

and

$$H = [-D_y^s/\epsilon, D_x^s/\epsilon, 0, B_y^s/\mu, -B_z^s/\mu, 0]^t. \tag{11}$$

As can be readily seen from the previous analytical statements, the medium is assumed to be isotropic, nondispersive and linear. Since the finite volume scheme is based on flux evaluations at cell interfaces, the vector $\hat{U}$ is comprised of the electric and magnetic fluxes $\mathbf{D}$ and $\mathbf{B}$, respectively. By choosing $\mathbf{B}$ over $\mathbf{H}$ (the magnetic intensity) and $\mathbf{D}$ over $\mathbf{E}$ (the electric field), we naturally preserve the continuity of normal components at material interfaces, where the fluxes are evaluated.

The adoption of the scattered field formulation requires that

$$J = [M_{dx}, M_{dy}, M_{dz}, J_{cx} + J_{dx}, J_{cy} + J_{dy}, J_{cz} + J_{dz}], \tag{12}$$

where the addition of the conduction current $\mathbf{J_c}$ and the polarization current of the dielectric $\mathbf{J_d}$ is given by

$$\mathbf{J}_c + \mathbf{J}_d = \frac{\sigma}{\epsilon}(\mathbf{D}^s + \mathbf{D}^i) + \left(1 - \frac{\mu_o}{\mu}\right)\frac{\partial \mathbf{D}^i}{\partial t}. \tag{13}$$

For the magnetization currents, we write:

$$\mathbf{M}_d = \left(1 - \frac{\epsilon_o}{\epsilon}\right)\frac{\partial \mathbf{B}^i}{\partial t}. \tag{14}$$

As expected, when the conductivity σ is zero and the space is nonmaterial, the conduction, polarization and magnetization currents assume a value of zero. For this situation, Maxwell's equations are devoid of the incident fields $\mathbf{D}^i$ and $\mathbf{B}^i$. However, if a perfectly conducting body is present, $\mathbf{D}^i$ reappears in the boundary condition that requires that the total electric field tangential to the perfect electrical conductor be zero. Similarly, the total magnetic field normal to the surface is likewise zero. In mathematical terms, we write,

$$\mathbf{n} \times (\mathbf{D}^i + \mathbf{D}^s) = 0 \qquad\qquad \mathbf{n} \cdot (\mathbf{B}^i + \mathbf{B}^s) = 0 \tag{15}$$

The discretization of the conservative equation is accomplished via a cell–centered, finite volume approach in conjunction with the Runge–Kutta, two–stage integrator. First consider the spatial discretization. By following the lead of Steger and Warming [8], we split the fluxes at any cell interface in terms of its positive and negative components (e.g., F^+ and F^-) as defined from the eigenvalue analysis of the conservative equation in the one–dimensional

time–space plane. These fluxes are associated with the right– and left–running waves at the cell interface; the states U^L and U^R are constructed from known values of U in adjacent cells. For example, the windward–biased scheme of Anderson et al. [9] requires that at interface $i + 1/2$,

$$U_{i+1/2}^L = U_i + \frac{\phi}{4}[(1 - \kappa)\nabla + (1 + \kappa)\Delta]U_i, \tag{16}$$

and

$$U_{i+1/2}^R = U_{i+1} - \frac{\phi}{4}[(1 + \kappa)\nabla + (1 - \kappa)\Delta]U_{i+1}, \tag{17}$$

where ϕ is a limiter ($\phi = 0, 1$), κ is an accuracy parameter, $\nabla U_i = U_i - U_{i-1}$ and $\Delta U_i = U_{i+1} - U_i$. When ϕ equals unity and κ takes on a value of $-1, 1/3$ or $+1$, the scheme is deemed second–order windward, third–order windward–biased or second–order central–differenced, respectively; the second–order Fromm scheme is recovered when $\phi = 1$ and $\kappa = 0$. Thus one can see that the advantage of the scheme of Anderson is found on its ability to tailor the scheme in a fashion that captures the specific physics of the problem at hand. Although third–order accuracy is desired in most situations, a second–order windward scheme has the advantage of predicting the slope of the field at a dielectric interface from field values totally resident in the dielectric. In contrast, a central difference approximation will lead to boundary errors since the prediction of the dependent variable requires knowledge of the dependent variable on both sides of the dielectric interface; for large disparities between the interfacial permittivities, the discontinuities in either normal **E** or tangential **D** will not be captured. To predict these discontinuities correctly, two options exist: 1) use a fully windward scheme that will introduce some oscillation in the data or 2) set ϕ to a value of zero. For this latter case, a certain amount of artificial dissipation will be introduced into the solution due to the resulting first–order approximation.

Once the left and right states of U are estimated at the $i + 1/2$ interface, the flux crossing that surface is simply,

$$F_{i+1/2} = F^+(U_{i+1/2}^L) + F^-(U_{i-1/2}^R). \tag{18}$$

Similarly, at interfaces j and k:

$$G_{j+1/2} = G^+(U_{j+1/2}^L) + G^-(U_{j-1/2}^R) \qquad (19)$$

and

$$H_{k+1/2} = H^+(U_{k+1/2}^L) + H^-(U_{k-1/2}^R). \qquad (20)$$

For generalized coordinates, the flux, say $\hat{F}$, is split in the direction of ξ. This is accomplished by a local transformation matrix T and by defining a new flux $\bar{F}$: $\bar{F} = T\hat{F}$. Since T can be chosen such that $\bar{F}$ and $\hat{F}$ have the same functional form, the eigenvalue/eigenvector information is directly obtained from the Cartesian formulation. Thus, if $\bar{F} = \bar{F}^+ + \bar{F}^-$ then $\hat{F} = \hat{F}^+ + \hat{F}^-$, where $\hat{F}^+ = T^{-1}\bar{F}^+$ and $\hat{F}^- = T^{-1}\bar{F}^-$. This same procedure is repeated in the directions η and ζ, for $\hat{G}$ and $\hat{H}$, respectively.

We now turn to the temporal integration. The two–stage Runge–Kutta (RK) integrator is algorithmically defined in terms of the following sequence:

$$U_1 = U^n + \delta t \frac{\partial U^n}{\partial t} \qquad (21)$$

$$U^{n+1} = .5(U^n + U_1) + .5\delta t \frac{\partial U_1}{\partial t}, \qquad (22)$$

where the derivative of U is known from the spatial discretization of (1). There are several attributes of the RK method that makes it attractive for this investigation: 1) The fully explicit nature of the scheme leads to codes that can be vectorized in a straight–forward fashion; 2) when dispersive media is present, the temporal nature of the media is represented by a simple inclusion of a polarization current; 3) if higher–order temporal accuracy is desired or required, the framework for the four–stage RK scheme is already in place; and 4) the boundary values of the dependent variables can be prescribed in a subroutine separate of the time integrator.

Fourier Analysis

The output of the FVTD code is time–domain field data. In order to achieve wide bandwidth responses, a wide–band input must be used. This is best accomplished by allowing

the incident field to be a gaussian pulsed plane wave. For example, let

$$E_x^{inc}(x,t) = e^{-w^2(t-(x-x_o)/c)^2},$$ (23)

where w controls the width of the pulse and x_o positions the pulse at the time $t = 0$.

To understand the wide–band nature of the pulse, consider the following Fourier pairs:

$$E_x^{inc}(x,t) = \int_{-\infty}^{\infty} e_x^{inc}(x,\omega)e^{j\omega t}d\omega$$ (24)

and

$$e_x^{inc}(x,\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e_x^{inc}(x,t)e^{-j\omega t}dt.$$ (25)

For the pulse described in (23), the Fourier transform is simply

$$e_x^{inc}(x,\omega) = \frac{1}{2w\sqrt{\pi}}e^{-\omega^2/(2w)^2}e^{-j\omega t_o},$$ (26)

where $t_o = (x - x_o)/c$; c is the speed of light. Note: as w becomes larger, the pulse becomes narrower and its frequency spectrum becomes broader; as w becomes smaller, the pulse becomes broader and the frequency spectrum becomes narrower. Thus, w is viewed as the parameter that controls the amount of frequency information in the input pulse. Although it is desirable to have wide–band inputs in theory, this is not possible in practice since the algorithm must discretize according to the smallest wavelength (or highest frequency).

To determine the frequency–domain response of one of the dependent variables, say E_x^s, a running Fourier transform is required. That is, the integral

$$e_x^s(\mathbf{x},\omega) = \int_0^t E_x^s(\mathbf{x},\tau)e^{j\omega\tau}d\tau$$ (27)

is calculated as a weighted sum of the time–domain response and is updated after each time step. A separate summation is required for each value of $\mathbf{x}$. The running summation is discontinued in time when the aggregate of terms no longer contributes to the final value of $e_x^s(\mathbf{x},\omega)$. The impulse response, which is the desired response for the RCS calculation is

$$\frac{e_x^s(\mathbf{x},\omega)}{e_x^{inc}(\mathbf{x},\omega).}$$ (28)

63–9

The previous expression is theoretically valid for all frequencies; unfortunately, there exists an upper frequency for which numerical noise will corrupt the information contained in the impulse response.

RCS Calculation

RCS data is readily available by means of the Schelkunoff equivalence principle of secondary sources and the Fourier transform. By defining a suitable, closed–integration surface S, we can deduce the power at any point exterior to S by means of the frequency domain formula [10]

$$P^s(R, \theta, \phi) = \frac{k^2 \eta}{2(4\pi R)^2} \left[\mathcal{A}_\theta \mathcal{A}_\theta^* + \mathcal{A}_\phi \mathcal{A}_\phi^* + \frac{1}{c^2} \left(\mathcal{F}_\theta \mathcal{A}_\theta^* + \mathcal{F}_\phi \mathcal{A}_\phi^* \right) + \frac{2}{c} \mathcal{R}e \left\{ \mathcal{A}_\theta \mathcal{F}_\phi^* + \mathcal{A}_\phi \mathcal{F}_\theta^* \right\} \right]$$

(29)

where, for example,

$$\mathcal{A}_\theta = \int_S [\cos\theta \cos\phi K_x + \cos\theta \sin\phi K_y - \sin\theta K_z] e^{k\mathcal{L}} dS \qquad (30)$$

and

$$\mathcal{F}_\theta = \int_S [\cos\theta \cos\phi K_{xm} + \cos\theta \sin\phi K_{ym} - \sin\theta K_{zm}] e^{k\mathcal{L}} dS. \qquad (31)$$

Here $\mathbf{K}$ and $\mathbf{K}_m$ are the equivalent, time–harmonic electric and magnetic surface currents given in terms of $\mathbf{n} \times \mathbf{h}^s$ and $-\mathbf{n} \times \mathbf{e}^s / \mu_o$, respectively; $\mathcal{L}$ is the radiation phase function. If P^i is the power in the incident pulse, then

$$\text{RCS} = \lim_{R \to \infty} 4\pi R^2 \frac{P^s}{P^i}. \qquad (32)$$

The computation of $\mathcal{A}$ and $\mathcal{F}$ is put into effect by discretizing the surface S into contiguous patches. Since the patch areas and their normals are computed previously for the flux calculations, the integration is reduced to a weighted sum of equivalent surface currents multiplied by the area and the radiation phase function.

RCS Results

The time–domain, characteristic–based method has been employed to predict the radar cross–section (RCS) of a perfectly conducting sphere. Consider Figure 1, which shows the
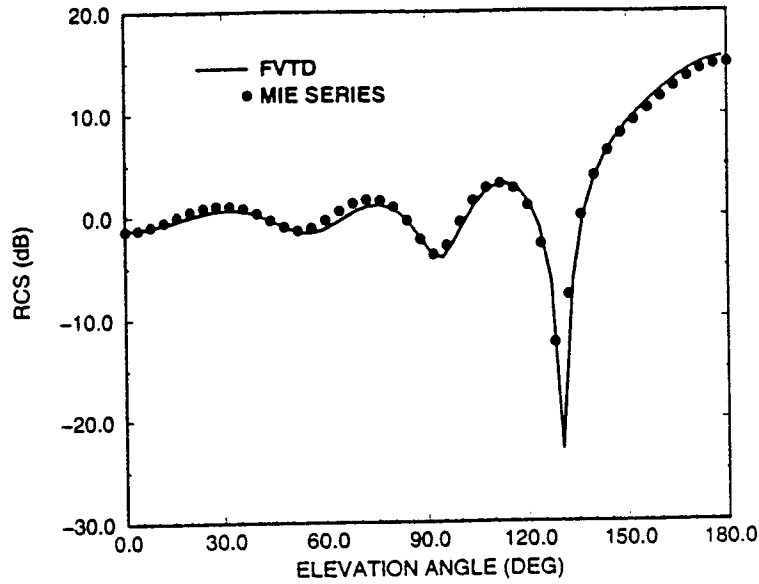
Figure 1: The RCS of a sphere; $ka = 5.3$, $\phi = 0°$.

RCS as a function of elevation angle when $\phi = 0°$; the electrical radius (i.e., ka) is .5.3. On this same plot is the solution obtained from the Mie series. Corresponding to the plane $\phi = 90°$, the RCS as a function of elevation angle is plotted against the Mie solution in Figure 2. As can be seen from both plots, the agreement is quite good.

The computational domain is spherical in nature such that I, J, K denote the number of grid lines in the R, θ, ϕ directions, respectively; Figure 3 shows a depiction of the grid; the core sphere is the perfectly conducting scatterer. For the present simulation, it suffices to set I, J, K to $73, 60, 96$. For a computational radius of 2 meters, a perfectly conducting sphere radius of .5 meters, a normalized phase velocity, and a CFL of unity, 6000 time steps are needed to capture the entire scattering event. The large number of time steps is required due to explicit integration, the Courant stability limit and the small cell volumes at the poles of the sphere.

Currently, the RCS of a dielectrically coated sphere is being studied. The purpose of this study is to see how well the time–domain finite–volume method predicts the late time
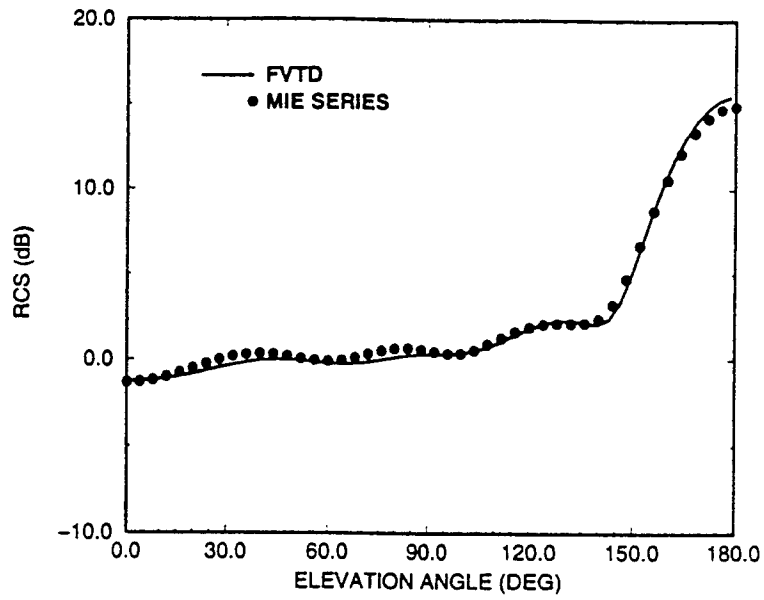
Figure 2: The RCS of a sphere; $ka = 5.3$, $\phi = 90°$.
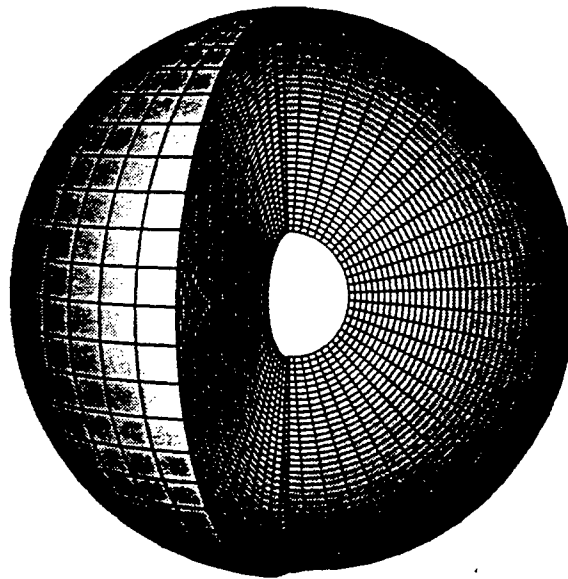
Figure 3: A cut–away section of the computational grid.

response of resonant–type structures. Verification of the data is currently being conducted.

New Boundary Condition for a PEC

Although the finite–volume scheme has been applied to a large class of problems and been shown to yield accurate results, it has accuracy limitations at perfectly conducting interfaces. Since all of the independent variables are collocated at cell centers and are not known on the boundary, an extrapolation scheme must be developed to estimate the various boundary values. This extrapolation and its implementation in the finite–volume algorithm will lead to errors.

As outlined below, an extrapolation scheme is not necessary. Instead, from Maxwell's equations it is possible to derive a set of governing equations for the surface charge and currents. Using this set of equations for the perfect conductor and Maxwell's equations for the interior, we can completely and exactly define the physics. The final accuracy of the scheme will be associated only with the discretization procedure.

At a perfectly conducting interface, $\mathbf{n} \times \mathbf{E} = 0$, $\mathbf{n} \times \mathbf{H} = \mathbf{J}_s$, $\mathbf{n} \cdot \mathbf{D} = \rho_s$ and $\mathbf{n} \cdot \mathbf{B} = 0$, where $\mathbf{J}_s$ is the induced surface current and ρ_s is the induced surface charge [11]. Unfortunately $\mathbf{J}_s$ and ρ_s are not known *a priori*, since their values are dependent on surface geometry and field excitation. This difficulty can be circumvented by reconsidering some special cases of Maxwell's equations:

$$\frac{\partial \mathbf{D}}{\partial t} = \nabla \times \mathbf{H} \tag{33}$$

and

$$\frac{\partial \mathbf{B}}{\partial t} = -\nabla \times \mathbf{E}. \tag{34}$$

In the development below, we are particularly interested in finding solutions of Maxwell's equations on the surface of a perfect conductor.

First, it is well known that Ampere's law is consistent with the law of charge conservation. For surface quantities, the consistency is readily seen by taking the dot product of Ampere's law with the surface's unit normal. After some vector manipulations and substitutions, one

finds

$$\frac{\partial \rho_s}{\partial t} + \nabla \cdot \mathbf{J}_s = 0; \quad \mathbf{x} \in \Gamma, \tag{35}$$

where Γ is the perfectly conducting surface and $\mathbf{x}$ is the position vector. In the previous equation, it is implicitly understood that the divergence operator is locally two–dimensional and is defined on the plane of the conductor.

A lesser–known relationship can be obtained from Faraday's law; this time, however, the cross–product is invoked. That is, by crossing $\mathbf{n}$ (the normal to the surface) with the left– and right–hand sides of Faraday's law, we obtain

$$\mathbf{n} \times \frac{\partial \mathbf{B}}{\partial t} = -\nabla(\mathbf{n} \cdot \mathbf{E}) + (\mathbf{n} \cdot \nabla)\mathbf{E}. \tag{36}$$

When $\mathbf{E}$ and ∇ are decomposed into their normal and tangential components (relative to the surface), Eqn. (36) takes on the equivalent form

$$\mathbf{n} \times \frac{\partial \mathbf{B}}{\partial t} = [-\mathbf{t} \cdot \nabla(\mathbf{E} \cdot \mathbf{n}) + \mathbf{n} \cdot \nabla(\mathbf{E} \cdot \mathbf{t})]\mathbf{t}, \tag{37}$$

where $\mathbf{t}$ is the tangential unit vector. The substitution of $\epsilon\mathbf{n} \cdot \mathbf{E} = \rho_s$ and $\mathbf{n} \times \mathbf{B} = \mu\mathbf{J}_s$ into Eqn. (37) yields

$$\mu\frac{\partial \mathbf{J}_s}{\partial t} = [-\mathbf{t} \cdot \nabla(\rho_s/\epsilon) + \mathbf{n} \cdot \nabla(\mathbf{E} \cdot \mathbf{t})]\mathbf{t}; \quad \mathbf{x} \in \Gamma. \tag{38}$$

Since it is understood that the current density flows tangential to the surface, it is appropriate to write for the previous expression,

$$\mu\frac{\partial \mathbf{J}_s}{\partial t} + \nabla(\rho_s/\epsilon) = \mathbf{F}; \quad \mathbf{x} \in \Gamma, \tag{39}$$

where $\mathbf{F}$ is the forcing function, defined by $\mathbf{F} = \mathbf{n} \cdot \nabla(\mathbf{E} \cdot \mathbf{t})\mathbf{t}$.

Equations (35) and (39) constitute the governing equations for the surface current and the surface charge associated with a perfectly conducting surface. Before a description of the full numerical procedure is given, a couple of observations are in order. First, the surface current and surface charge satisfy the acoustic–type equations. This observation is not unanticipated, since current is analogous to fluid velocity and charge density is analogous to mass density. Second, the circulation of the current is proportional to the circulation of

the normal derivative of the tangential electric field, which is the source term of the system. For this reason the current/charge wave will be a linear combination of longitudinal and transverse waves. Finally, in terms of a numerical procedure, the source term is computable, since the tangential electric field is zero on the conductor and is known in the domain from Maxwell's equations.

To summarize, let Γ define the perfectly conducting boundary and let Ω define the computational domain (less boundary). Then for fields in Ω,

$$\frac{\partial \mathbf{D}}{\partial t} - \nabla \times \mathbf{H} = 0 \tag{40}$$

and

$$\frac{\partial \mathbf{B}}{\partial t} + \nabla \times \mathbf{E} = 0; \tag{41}$$

for fields on the boundary,

$$\frac{\partial \rho_s}{\partial t} + \nabla \cdot \mathbf{J}_s = 0 \tag{42}$$

and

$$\mu \frac{\partial \mathbf{J}_s}{\partial t} + \nabla (\rho_s/\epsilon) = \mathbf{F}. \tag{43}$$

These two systems are coupled at the boundary by means of $\mathbf{n} \times \mathbf{E} = 0$, $\mathbf{n} \times \mathbf{H} = \mathbf{J}_s$, $\mathbf{n} \cdot \mathbf{D} = \rho_s$ and $\mathbf{n} \cdot \mathbf{B} = 0$.

Paper Submission

The work performed this past summer is being continued as a collaborative effort with Dr. Y. Weber. The focus of this investigation is on the two–dimensional scattering from a perfectly conducting channel. An abstract is being prepared for submission to the AIAA conference in New Orleans, summer of 1996. Dr. Weber will be the lead author; Dr. K. Hill (Wright Patterson Air Force Base) Dr. J. L. Young are secondary authors.

References

[1] Yee, K. S., "Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media," *IEEE Trans. Ant. Propagat.*, vol. 14, pp. 302–307, 1966.

[2] Shankar, V., A. H. Mohammadian and W. F. Hall, "A time–domain, finite-volume treatment for the Maxwell equations," *Electromagnetics*, vol. 10, pp. 127–145, 1990.

[3] Mahadevan, K., and R. Mittra, "Radar cross section computation of inhomogeneous scatterers using edge–based finite element methods in the frequency and time domains," *Radio Science*, vol. 28, pp. 1181–1193, 1993.

[4] Madsen, N. K., "Divergence preserving discrete surface integral methods for Maxwell's curl equations using non–orthogonal unstructured grids," *Research Institute for Advanced Computer Science* Tech. Rep. 92.04, NASA Ames Research Center, 1992.

[5] Shang, J. S. and D. Gaitonde, "Characteristic-based, time–dependent Maxwell equations solvers on a general curvilinear frame," *AIAA Journal*, vol. 33, pp. 491–498, 1995.

[6] Shang, J. S., and D. Gaitonde, "Scattered electromagnetic field of a re–entry vehicle," *J. Spacecraft & Rockets*, vol. 32, pp. 294–301, 1995.

[7] Shang, J. S., "Characteristic–based algorithms for solving the Maxwell equations in the time–domain," *IEEE Ant. Propagat. Mag.*, vol. 37, pp. 15–25, 1995.

[8] Steger, J. L., and R. F. Warming, "Flux vector splitting of the inviscid gasdynamic equations with application to finite–difference methods," *J. Comp. Phys.*, vol. 40, pp. 263–293, 1981.

[9] Anderson, W. K., J. L. Thomas and B. van Leer, "A comparison of finite volume flux vector splittings for the Euler equations," *AIAA 23rd Aerospace Sciences Meeting*, Reno, NV, AIAA 85–0122, 1985.

[10] Elliot, R. S., *Antenna Theory and Design*, Prentice–Hall, New York, 1981.

[11] Harrington, R. F., *Time–Harmonic Electromagnetic Fields*, McGraw–Hill, New York, 1961.

Acknowledgements

MISSILE AUTOPILOT DESIGN
BASED ON A UNIFIED SPECTRAL THEORY
FOR LINEAR TIME-VARYING SYSTEMS

J. Jim Zhu
Assistant Professor
Department of Electrical and Computer Engineering

with

Michael C. Mickle
Candidate for Doctor of Philosophy
Department of Electrical and Computer Engineering

Louisiana State University
Baton Rouge, LA 70803
Email: zhu@sun-ra.rsip.lsu.edu

Final Report for:
Summer Faculty Research Program
Wright Laboratory

July 1995

MISSILE AUTOPILOT DESIGN
BASED ON A UNIFIED SPECTRAL THEORY
FOR LINEAR TIME-VARYING SYSTEMS

J. Jim Zhu
Assistant Professor
Department of Electrical and Computer Engineering
Louisiana State University

Abstract

This report presents design and simulation case studies of a missile autopilot for angle of attack and nominal acceleration tracking using a recently developed Extended-Mean Assignment (EMA) control technique for linear time-varying (LTV) systems. The EMA control technique is based on a new eigenvalue concept, called SD-eigenvalue, for LTV systems. Closed-loop stability is achieved by the assignment of the extended-mean of these time-varying SD-eigenvalues to the left-half complex plane in a way similar to the eigenvalue (pole) assignment technique for linear time invariant (LTI) systems. Salient features of the tracking controller include: (i) good tracking performance for arbitrary trajectories without any scheduling of the constant design parameters throughout the entire operating range of the Mach, (ii) implementation of the inverse pitch dynamics using a static neural network, (iii) time-varying EMA control gains to improve tracking performance, and (vi) a time-varying bandwidth command shaping filter that effectively reduces the actuator rate while maintaining good tracking response for both smooth and abrupt trajectories. Although the autopilot was designed only for nominal aerodynamic coefficients and constant trajectories, excellent performance was verified for $\pm 50\%$ variations in the aerodynamic coefficients, and for arbitrary command trajectories.

MISSILE AUTOPILOT DESIGN
BASED ON A UNIFIED SPECTRAL THEORY
FOR LINEAR TIME-VARYING SYSTEMS

J. Jim Zhu

1. Introduction

This report presents design and simulation case studies of a missile autopilot design using a recently developed Extended-Mean Assignment (EMA) control technique for linear time-varying (LTV) systems. The EMA control technique is very similar to the conventional pole placement design for linear time-invariant (LTI) systems, but based on a new, time-varying Series D-eigenvalue (SD-eigenvalue) concept [1]-[4]. The autopilot is to control the nonlinear time-varying pitch-axis dynamics of a hypothetical tail-controlled missile, which has been used as a benchmark in a number of recent studies on nonlinear gain-scheduling design techniques [5]-[7]. In [1] the theory and design technique were presented for an EMA controller that uses complex-valued SD-eigenvalues to avoid singularities known as finite-escapes. The nonlinear pitch dynamics of the missile was rendered into a linear time-varying system via the classical linearization along a nominal trajectory, and then operated on by a linear coordinate transformation to make it tractable by the EMA control technique. Simulation validation of the zero input stability was also presented there.

In this report, we present the design and simulation study of trajectory tracking using the complex-valued EMA controller. A radical departure from the conventional design philosophy is that nonlinearity and time-variance are not treated as nuisances, but purposely utilized to accomplish design objectives beyond the capability of LTI controllers. Salient features of the EMA tracking controller include: (i) good tracking performance for arbitrary trajectories without any scheduling of the constant design parameters throughout the entire operating range of the Mach, (ii) implementation of the inverse pitch dynamics using a static neural network, (iii) time-varying EMA control gains to improve tracking performance, and (vi) a time-varying bandwidth command shaping filter that effectively reduces the actuator rate while maintaining good tracking response for both smooth and abrupt trajectories.

For completeness, Section 2 recapitulates the main results presented in [1]. Section 3 details the design and implementation of the EMA tracking controller, including: the Radial Basis Function (RBF) neural network based inverse plant model, the time-varying EMA command logic, and the time-varying bandwidth command shaping filter. Two Normal Acceleration (NA) tracking system configurations are designed; both are centered around an Angle of Attack (AOA) tracking subsystem. One configuration uses an AOA state observer to estimate the AOA tracking error from that of the NA measurement. The other uses the AOA subsystem as an inner loop, and employs a Proportional-Integral (PI) controller for the NA outer loop tracking. In Section 4, simulation case studies are presented for: (i) AOA tracking of step trajectories with constant EMA commands, and with nominal

and $\pm 50\%$ variations in the aerodynamic coefficients, (ii) AOA tracking of sinusoidal trajectories with both constant and variable EMA commands, and with nominal and $\pm 50\%$ variations in the aerodynamic coefficients, (iii) NA tracking of both step and sinusoidal trajectories using an AOA state observer, and (iv) NA tracking of both step and sinusoidal trajectories using AOA inner loop. Section 5 concludes this report with a summary of the results and suggestions for further studies on tracking of arbitrary normal acceleration trajectories using a dynamic neural network based "pseudo-inverse" of the non-minimum phase plant model.

2. Preliminaries

This section summarizes the main results of [1]. Interested readers are referred to [1] for detailed derivations and the missile model parameters. It is noted that some typos found in [1] are corrected here.

2.1 Complex-valued EMA controller

The EMA control technique is based on a new SD-eigenvalue concept for LTV systems in a way similar to the conventional pole placement design method for LTI systems. Let $D = d/dt$ be the derivative operator. A 2nd-order LTV system

$$\ddot{y} + \alpha_2(t)\dot{y} + \alpha_1(t)y = u \tag{2.1}$$

can be written in an operator form $\mathcal{D}_\alpha\{y\} = u$ where

$$\begin{aligned}\mathcal{D}_\alpha &= D^2 + \alpha_2(t)D + \alpha_1(t) \\ &= (D - \lambda_2(t))(D - \lambda_1(t))\end{aligned} \tag{2.2}$$

is known as a *polynomial differential operator* (PDO), and the factorization is known as *Cauchy-Floquet factorization*. The scalar functions $\lambda_1(t), \lambda_2(t)$ are called *Series D-eigenvalues* (*SD-eigenvalues*) for the LTV system (2.1). The SD-eigenvalues satisfy a set of (nonlinear, differential) SD-characteristic equations

$$\begin{aligned}\dot{\lambda}_1(t) + \lambda_1^2(t) + \alpha_2(t)\lambda_1(t) + \alpha_1(t) &= 0 \\ \lambda_2(t) &= -\alpha_2(t) - \lambda_1(t)\end{aligned} \tag{2.3}$$

Note that, in general, $\lambda_1(t)$, $\lambda_2(t)$ are non-commutative, nonunique, and may be complex-valued. In the latter case, they form an *affine complex-conjugate* pair [6]

$$\begin{aligned}\lambda_1(t) &= \sigma_1(t) + j\omega(t) \\ \lambda_2(t) &= \sigma_2(t) - j\omega(t)\end{aligned} \tag{2.4}$$

where $\omega(t)$ satisfies

$$\dot{\omega} = (\sigma_2(t) - \sigma_1(t))\omega \tag{2.5}$$

Now define the extended-mean (EM) value of an integrable function $\lambda(t)$ by

$$\mathrm{em}\{\lambda(t)\} = \limsup_{(t-t_0)\to\infty} \frac{1}{t - t_0} \int_{t_0}^{t} \lambda(\tau)d\tau \tag{2.6}$$

Then the LTV system is exponentially stable if the EM values of $\lambda_1(t)$, $\lambda_2(t)$ are in the LHP of $\mathbb{C}$, *i.e.*

$$\mathrm{em}\{\mathrm{Re}(\lambda_i(t))\} < 0, \qquad i = 1, 2 \tag{2.7}$$

Thus, if the LTV system (2.1) is unstable, a feedback control law

$$u(t) = k_1(t)y(t) + k_2(t)\dot{y}(t) \tag{2.8}$$

can be synthesized so that SD-eigenvalues $\gamma_1(t)$, $\gamma_2(t)$ of the closed-loop system $\mathcal{D}_\eta\{y\} = 0$ where

$$\mathcal{D}_\eta = D^2 + \eta_2(t)D + \eta_1(t) \qquad (2.9)$$
$$= (D - \gamma_2(t))(D - \gamma_1(t))$$

with

$$\eta_i = \alpha_i(t) - k_i(t)$$

has the desired EM values $C_i(t)$. These desired EM values can be achieved using LTI tracking control methods to drive the EMA error

$$\epsilon_i(t) = \text{em}\{\gamma_i(t)\} - C_i(t) \to 0 \qquad (2.10)$$

exponentially, thereby rendering a LTV control problem to a LTI one.

A nuisance in the previously developed EMA controller is that the SD-eigenvalues may have finite singularities known as finite escapes. Even though some methods have been developed to circumvent the problem, they are either restricted to 2nd-order systems only, or merely render the infinite values in $\lambda_i(t)$ to finite peaks that will eventually show up in the controlled motion.

In a recent paper [2], it is proven that for LTV systems with piecewise smooth coefficients, there always exist piecewise smooth SD-eigenvalues free of finite escapes. However, these well-behaved SD-eigenvalues are often complex-valued. A condition for the SD-eigenvalues of a 2nd-order LTV system to be well behaved is to allow complex-valued solutions

$$\lambda_{1,2}(t) = \sigma_{1,2}(t) \pm j\omega(t) \qquad (2.11)$$

for (2.3) when the discriminant

$$\theta = 2\dot{\alpha}(t) + \alpha_2^2(t) - 4\alpha_1(t) \le 0 \qquad (2.12)$$

Otherwise, let

$$\lambda_{1,2}(t) = \sigma_{1,2}(t) \pm \omega(t) \qquad (2.13)$$

By substituting (2.11)-(2.13) into (2.3), the SD-characteristic equation becomes:

$$\dot{\sigma}_1 + \sigma_1^2 + \alpha_2(t)\sigma_1 + \alpha_1 + \text{sgn}(\theta)\omega^2 = 0$$
$$\sigma_2 = -\alpha_2(t) - \sigma_1$$
$$\dot{\omega} = (\sigma_2(t) - \sigma_1(t))\omega \qquad (2.14)$$

These equations give the following useful relationship

$$\alpha_1(t) = -\dot{\sigma}_1(t) + \sigma_1(t)\sigma_2(t) - \text{sgn}(\theta)\omega^2(t)$$
$$\alpha_2(t) = -(\sigma_1(t) + \sigma_2(t)) \qquad (2.15)$$

Since only the EM values of the real-parts of the closed-loop SD-eigenvalues $\gamma_i(t)$ determine stability, we can still use real-valued EM compensating signals $\mu_i(t)$. This results in complex-valued closed-loop SD-eigenvalues $\gamma_i(t) = \psi_i(t) \pm j\phi(t)$, which prevents finite-escapes in the closed-loop systems, and allows design freedom in closed-loop dynamic behavior. Thus, we have

$$\gamma_i(t) = (\sigma_i(t) + \mu_i(t)) \pm j\phi(t) \qquad (2.16)$$
$$= \psi_i(t) \pm j\phi(t)$$

where $\phi(t)$ satisfies

$$\dot{\phi} = (\psi_2(t) - \psi_1(t))\phi \qquad (2.17)$$

Note that the same relationship (2.15) holds for $\eta_i(t)$ and $\gamma_i(t)$. Thus, the LTV feedback control gains $k_i(t)$ are given by

$$k_1 = \alpha_1 - \eta_1 = \dot{\mu}_1 - \sigma_1\mu_2 - \sigma_2\mu_1 - \mu_1\mu_2 - \text{sgn}(\theta)\omega^2 - \phi^2$$

$$k_2 = \mu_1 + \mu_2 \qquad (2.18)$$

where the sign for $\phi^2(t)$ has been fixed to obtain damped oscillatory impulse response for the closed loop system. The overall implementation diagram for the complex-valued EMA controller is given in Figure 1. Interested readers should compare it to the original EMA controller given in [4].

2.2 Dynamic model of the missile airframe

We now consider a hypothetical tail-controlled missile whose pitch-axis dynamics are described by

$$\dot{\alpha}(t) = K_\alpha M(t) C_n[\alpha(t), \delta(t), M(t)]\cos(\alpha(t)) + q(t)$$

$$\dot{q}(t) = K_q M^2(t) C_m[\alpha(t), \delta(t), M(t)]$$

$$\eta_z(t) = K_z M^2(t) C_n[\alpha(t), q(t), \delta(t), M(t)] \qquad (2.19)$$

where $\alpha(t)$ is the angle-of-attack (AOA), $q(t)$ is the pitch-rate, $\delta(t)$ is the tail-fin deflection angle, $\eta_z(t)$ is the normal acceleration (NA), and $M(t)$ is the Mach number. The aerodynamic lift coefficient C_n and the pitch moment coefficient C_m are modeled by

$$C_n[\alpha, \delta, M] = a_n\alpha^3 + b_n\alpha|\alpha| + c_n(2 - \frac{M}{3})\alpha + d_n\delta$$

$$C_m[\alpha, q, \delta, M] = a_m\alpha^3 + b_m\alpha|\alpha| + c_m(-7 + \frac{8M}{3})\alpha + d_m\delta \qquad (2.20)$$

The tail-fin actuator dynamics is described by

$$\frac{d}{dt}\begin{bmatrix} \delta(t) \\ \dot{\delta}(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\omega_a^2 & -2\zeta\omega_a \end{bmatrix}\begin{bmatrix} \delta(t) \\ \dot{\delta}(t) \end{bmatrix} + \begin{bmatrix} 0 \\ \omega_a^2 \end{bmatrix}\delta_c(t) \qquad (2.21)$$

where $\delta_c(t)$ is the commanded tail-fin deflection angle.

The values of the various constant parameters in the dynamic equations (2.19)-(2.21) can be found in [1] and [5]-[7], so they are omitted here. This missile model has been used as a benchmark in a number of recent studies on nonlinear gain-scheduling design techniques [5]-[7]. It is noted that in general the pitch moment coefficient C_m depends on the pitch rate q. This dependence is absent in the model (2.19) in the previous studies, and in this study as well. However, since for tail fin controlled pitch dynamics this dependence has a stabilizing effect, the performance should be improved when a controller designed with the model (2.19) is applied to the more accurate model.

Let $\eta_c(t)$ be the commanded NA trajectory, $\bar{\alpha}(t)$ be the corresponding nominal AOA trajectory, and $\bar{\delta}(t)$ be the required nominal control input. By applying the standard linearization along a nominal trajectory, followed by a linear coordinate transformation, a linear time-varying tracking error model was derived in [1] as follows

$$\ddot{z}_1 + \alpha_2(t)\dot{z}_1 + \alpha_1(t)z_1 = u \qquad (2.22a)$$

$$b_1(t)\dot{v} + [\dot{b}_1 + b_2(t)]v = u \tag{2.22b}$$

$$y = c(t)z_1 + d(t)v \tag{2.22c}$$

where

$$z_1(t) = \alpha(t) - \overline{\alpha}(t) \tag{2.23}$$

$$y(t) = \eta_z(t) - \eta_c(t) \tag{2.24}$$

$$v(t) = \delta(t) - \overline{\delta}(t) \tag{2.25}$$

and

$$\alpha_1(t) = -[\dot{a}_{11}(t) + a_{21}(t)] \tag{2.26}$$

$$\alpha_2(t) = -a_{11}(t) \tag{2.27}$$

$$b_1(t) = K_\alpha M(t)d_n\cos(\overline{\xi}_1(t)) \tag{2.28}$$

$$b_2(t) = K_q d_m M^2(t) \tag{2.29}$$

$$c(t) = K_z M^2(t)[3a_n\overline{\xi}_1^2(t) + 2b_n|\overline{\xi}_1(t)| + c_n(2 - \frac{M(t)}{3})] \tag{2.30}$$

$$d(t) = d_n K_z M^2(t) \tag{2.31}$$

where

$$a_{11}(t) = K_\alpha M(t)[(3a_n\overline{\xi}_1^2(t) + 2b_n|\overline{\xi}_1(t)| + c_n(2 - \frac{M(t)}{3}))\cos(\overline{\xi}_1(t)) -$$
$$- (a_n\overline{\xi}_1^3(t) + b_n|\overline{\xi}_1(t)|\overline{\xi}_1(t) + c_n(2 - \frac{M(t)}{3})\overline{\xi}_1(t) + d_n\overline{\delta}(t))\sin(\overline{\xi}_1(t))] \tag{2.32}$$

$$a_{21}(t) = K_q M^2(t)[3a_m\overline{\xi}_1^2(t) + 2b_m|\overline{\xi}_1(t)| + c_m(-7 + \frac{8M(t)}{3})] \tag{2.33}$$

This model is readily tractable by the EMA control technique, and will be used for the subsequent angle-of-attack tracking and normal acceleration tracking studies.

3. EMA Autopilot Design

To take advantage of the minimum phase zero-dynamics in AOA, and the LTV model (2.22) for AOA dynamics that is readily controlled by the EMA technique, the (nonminimum phase) normal acceleration (NA) tracking is achieved via an AOA tracking subsystem shown in Figure 3.1. Based on this AOA tracking subsystem, two NA tracking system configurations are proposed: one uses an observer to estimate the AOA tracking error from the NA tracking error, as shown in Figure 3.2; the other uses a PI controller to generate the AOA tracking command from the NA tracking error, as shown in Figure 3.3. The design of subsystem components are described below.

3.1 Time-Varying Bandwidth (TVB) Command Shaping Filter

Due to the stringent requirements on tracking performance and actuator rate limit, all three system configurations in Figures 3.1-3.3 require a tracking command shaping filter, which is not shown in the figures. The filter should greatly reduce the acceleration and rate of an abrupt command trajectory, whereas it should have little effect on smooth trajectories that can be tracked within the actuator limits. These two requirements cannot be achieved with a fixed-parameter filter. For instance, the AOA tracking subsystem can track a 0.5 Hz, 0.3 radians

amplitude sine wave within the ± 8.7 rad/sec (500°/sec) actuator rate limits without any command shaping. However, the same system tracks a step command of 0.3 radians manitude acurately, but requires a maximum actuator rate of 17,500 rad/sec. When a 3rd-order LTI Bessel filter with a bandwidth $\omega_n = 10$ is applied to the step tracking command, the actuator rate is reduced to within the limits with a satisfactory tracking performance, but the sinusoidal tracking then has an unnecessary tracking delay of 0.25 second, or a 45° phase lag.

To resolve this problem, a novel 2nd-order, LTV filter with a *variable bandwidth* is designed. The impulse response of the filter is chosen to be of the form

$$h(t) = A\exp(-\zeta \int \omega_n(t)dt)\cos(\int \omega_d(t)dt + \phi) \tag{3.1}$$

where ζ is a constant damping coefficient, ϕ is a constant depending on ζ, and $\omega_d(t)$ is related to $\omega_n(t)$ by

$$\omega_d(t) = \omega_n(t)\sqrt{1 - \zeta^2}$$

For a constant ω_n, this is the familiar 2nd-order LTI filter that has a damped oscillatory impulse response for $0 < \zeta < 1$. The dynamical equation of the filter is given by

$$\ddot{c}_{out} + a_2(t)\dot{c}_{out} + a_1(t)c_{out} = c_{in} \tag{3.2}$$

where c_{in} and c_{out} are the command input and the shaped command output. The time-varying system coefficients $a_1(t)$, $a_2(t)$ and the associated Parallel D-eigenvalues (PD-eigenvalues) $\rho_i(t)$ are given by [3]

$$a_1(t) = \omega_n^2(t) \tag{3.3}$$

$$a_2(t) = -\left[2\zeta\omega_n(t) + \frac{\dot{\omega}_n(t)}{\omega_n(t)}\right] \tag{3.4}$$

$$\rho_{1,2}(t) = -\zeta\omega_n(t) \pm j\omega_d(t) \tag{3.5}$$

Thus, the filter is well-defined if $\omega_n(t) > 0$, and is exponentially and BIBO stable if and only if, in addition to $\omega_n(t) > 0$, $\zeta > 0$. Incidentally, it is interesting to note that under this stability condition, the filter can be stable while $a_2(t)$ is negative when $\omega_n(t)$ decreases at a fast rate, which implies that a "frozen-time" eigenvalue is in the right-half-plane of $\mathbb{C}$!

For the application at hand, the damping factor $\zeta = 1$ is chosen, and the filter bandwidth $\omega_n(t)$ and its rate $\dot{\omega}_n(t)$ are generated from

$$\ddot{\omega}_n(t) + 2\zeta_0\omega_0\dot{\omega}_n(t) + \omega_0^2\omega_n(t) = \omega_0^2 r_\omega(t) \tag{3.6}$$

where ζ_0, ω_0 are design parameters that determine how $\omega_n(t)$ follows the bandwidth command $r_\omega(t)$, which is in turn determined by the following command shaping logic:

$$r_\omega(t) = \omega_{n0} - a \cdot \text{sat}(b \cdot \text{ddzone}(d \cdot |\ddot{c}_{out}(t)|)) \tag{3.7}$$

where ω_{n0} is the maximum bandwidth, a, b, d are constant design parameters which determine how the bandwidth is reduced from the maximum ω_{n0} when $|\dot{c}_{out}(t)|$ exceeds a predefined threshold set by the saturation function sat$(\cdot)$ and deadzone function ddzone$(\cdot)$, which are defined in the usual way.

Figures 3.7, 3.8 show the step and sine responses, respectively, of the TVB filter with a maximum $\omega_n(t) = \omega_{n0} = 100$, and minimum $\omega_n(t) = 5$, along with the responses of a 2nd-order LTI filter with a fixed $\omega_n = 100$ and a 3rd-order Bessel filter with a fixed $\omega_n = 10$. The corresponding filter parameters $a_1(t)$ and $a_2(t)$ are given in Figures 3.9, 3.10, respectively, which clearly show how the bandwidth is drastically reduced during the initial transient of an applied command trajectory.

3.2 Neural Network Based Dynamic Inverse

Nonlinear tracking by linearization along a nominal trajectory calls for an inverse model of the plant input/output dynamics to generate the required nominal control input. It is well known that when the plant has nonminimum phase zero-dynamics, such as the case of NA tracking, the inverse model is unstable. Consequently, "perfect tracking" is not possible. However, if the plant is stable, it is possible to find the inverse for constant input and output trajectories. The error in the nominal control for a variable trajectory is then to be compensated for by the tracking error controller, provided that the error is sufficiently small so that the linearization remains valid.

The nominal control for a constant trajectory is implemented by a Radial Basis Function (RBF) neural network (NN). Two static RBF NNs are trained, one generates the nominal fin deflection $\bar{\delta}$ for nominal AOA $\bar{\alpha}$, and the other generates both the nominal fin deflection $\bar{\delta}$ and nominal AOA $\bar{\alpha}$ for nominal NA. The former is used in the systems shown in Figures 3.1 and 3.3, and the latter is used in the system shown in Figure 3.2. The training data for the network was acquired from the MATLAB function trim which locates the equilibrium points of the missile model, *i.e.* the nominal tail fin deflection required to achieve the desired output (AOA or NA). These data were then used to train a RBF NN via the MATLAB function solverb. The training of a RBF NN with some 200 neurons required an order of 10^{10} FLOPs using the Unix version Neural Networks Toolbox Version 1.0 for MATLAB. Shown in Figures 3.4 and 3.5 are the desired inverse mapping of $\bar{\alpha} \mapsto \bar{\delta}$ and a 200-neuron RBF NN implementation, respectively. The error surface between these two is plotted in Figure 3.6, where it can seen that, except at a few peripheral points outside the operating range, the errors are below 0.01 (radians). This error magnitude is readily accommodated by the EMA tracking controller.

The RBF NN offers several advantages over other NNs. In general, it provides quicker training than networks such as the Multi-layer Network and does not have the problem of local minima. It provides smooth generalization between known data points, as opposed to the zeroth-order generalization of the CMAC. Also, it requires fewer neurons than CMAC. However, it shares CMAC's problem of exponential growth of the number of weights with respect to number of inputs, and is thus limited as a practical solution to problems with smaller number of inputs. It is noted that in practice, training data can be obtained directly from wind tunnel or flight tests. Although it takes significant amounts of time and computations for a comprehensive training, the localization of its receptive fields and the small number of weights makes the RBF NN ideal for on-line training, using the off-line training as the

initial states. This will greatly increase the accuracy and robustness of the overall tracking system in the presence of parameter uncertainties. These facts make the RBF well suited to the problem of generating a constant nominal control.

3.3 AOA Tracking Subsystem

The center piece and the novelty of the AOA tracking subsystem is the complex-valued EMA controller whose design was detailed in [1] and recapitulated in Section 2 of this report. It is noted that the design was greatly simplified owing to the fact that the AOA zero-dynamics (2.22b) is stable, as verified in [1] using the extended-mean stability criterion on SD-eigenvalues.

As the closed-loop stability is guaranteed by the extended-mean stability criterion, the extended-mean assignment commands (EMAC) $c_i(t)$ in the EMA controller need not be constant, as long as in the average they stay in the LHP of $\mathbb{C}$. This salient feature may be advantageous in cases where control energy is a prime concern while performance may be sacrificed during noncritical maneuverings. To test this concept, a variable EMAC logic is defined as follows

$$c_i(t) = c_{i0} + g\,\frac{\epsilon(t)}{\epsilon_0}\exp\left(-\frac{\epsilon(t)}{\epsilon_0}\right) \tag{3.8}$$

where $\epsilon(t)$ is the tracking error, ϵ_0 is a predetermined tracking error threshold, c_{i0} is the minimum EMAC, and g is a design constant that, when added to c_{i0}, determines the maximum EMAC. This EMAC logic generates the maximum EMAC when the tracking error $\epsilon(t)$ is equal to ϵ_0, and reduces to the minimum EMAC when $\epsilon(t)$ is either zero or infinity. This EMAC logic was implemented and tested in the simulation studies presented below.

3.4 AOA Observer for NA Tracking System

The missile guidance system typically generates a NA command profile. Thus, despite the suitability of the AOA dynamic model (2.22) for EMA control, and its nonminimum phase zero dynamics, it is often desirable to track the NA. It is, therefore, important to translate the AOA tracking strategy into a NA tracking controller. Since the EMA controller was designed to eliminate the tracking error in AOA, and NA is related to AOA and the tail fin deflection via a nonlinear algebraic mapping, it is natural to design a nonlinear time-varying state observer that estimates the AOA error from the measurement of the NA error. This is accomplished by inverting the linearized output error equation (2.22c). A lowpass filtered differentiator is used to estimate the AOA error rate. It is noted that the observer performance could be improved by using a static neural network to approximate the nonlinear relation between the NA and AOA errors. Other ways of implementing the observer were discussed in [1].

It is noted that this NA tracking strategy circumvents the nonminimum phase problem of NA tracking implicitly, and yields good results when the mapping between the NA and AOA errors is accurate. This method is

applicable to other nonminimum phase tracking problems, and may be termed *algebraic output redefinition*, as opposed to the (dynamic) output redefinition method proposed in [8].

3.5 AOA Inner-Loop for NA Tracking System

A second normal acceleration tracking structure currently being studied uses the AOA tracking subsystem as an inner loop for NA tracking, and employs a PI controller for the NA outer loop. The advantage of this strategy is that the integrated NA error would allow for zero steady state error to a step command. However, this option still requires further research to justify the outer-loop stability, and to improve transient performance deterioration due to the integral control. Only preliminary simulation results are shown below to exemplify the idea and problems.

4. Simulation Case Studies

Simulation studies were performed to validate the design. The TVB command shaping filter was used for all the cases, so that no actuator amplitude or rate limiter was used. The constant design parameters used in the EMA controller were fixed for all the cases at $k_{i0} = -100$, $k_{i1} = 0$, $k_{i2} = -10$.

Case 1: AOA Step Trajectory Tracking

The AOA EMA control provides remarkable results for step command tracking. Figure 4.1 displays a three-second piecewise constant AOA tracking command, the TVB filtered command, and the AOA output. Figure 4.2 shows the corresponding tail fin deflection rate which is well within the 8.7 rad/sec design constraint. Without the TVB filter, the EMA controller accurately tracks step commands, but the actuator rate reaches 17,000 rad/sec. Thus, the TVB filter is essential for limiting the actuator to achievable rates. The output tail deflection is also limited but as figure 4.3 indicates this was an inconsequential constraint. Figure 4.4 shows the results of simulations of the four possible combinations of ±50% variations in the two aerodynamic coefficients $C_n(t), C_m(t)$. Clearly, the AOA still accurately tracks the desired trajectory for all four cases, indicating excellent robustness of the closed-loop system.

Cease 2: AOA Variable Trajectory Tracking

Although the neural network was trained to generate a nominal control input only for static or step commands, the proposed controller configuration can track arbitrary trajectories because of the EMA section's ability to accommodate errors in the nominal control input. Also, the EMA assignment command need not be a constant. Thus, figures 4.5-4.11 compare the results for TVB filtered sinusoidal tracking with both constant and variable EMA command. Shown together in figure 4.5 are: the AOA tracking command, the TVB filtered command, the AOA output with constant EMA command at 20, and with variable EMA command between $10 \sim 20$ as defined in (3.8). It clearly shows the remarkable tracking performance. The filtered command has very little magnitude dampening and phase change, demonstrating how little the effect of the TVB filter has on a smooth trajectory comparing to its effect on the step command in the preceding case. Figure 4.7 shows the usefulness of the command filter for minimizing control rate and Figure 4.8 shows the corresponding fin deflection.

Shown in Figure 4.6 are the constant and variable EMA commands (EMAC). It can be seen that the variable EMAC indeed reduces to the minimum level of -10 when there is need for strenuous control action. Figures 4.9 and 4.10 show the corresponding feedback gains $k_1(t)$, $k_2(t)$, respectively, for both constant and variable EMAC. The tracking performance under constant and variable EMAC are almost indistinguishable in Figure 4.5, but Figure 4.11 indicates that in a long run the latter indeed saves control energy. .

As above, the robustness of the closed-loop system were tested for both constant and variable EMA commands under all four possible combinations of $\pm 50\%$ error in the two aerodynamic coefficients $C_n(t)$, $C_m(t)$. The results are shown in figure 4.12 which contains the eight test outputs together with the commanded trajectory. Once again, these results are very good.

Finally, the time-varying coefficients $\alpha_1(t)$, $\alpha_2(t)$ in the linearized AOA error dynamics (2.22), and the Mach profile as an important source of the time varying coefficients in the pitch airframe model, are shown in Figures 4.13-4.15 for the constant EMA command simulation. It is remarkable that no constant design parameters need to be scheduled for the entire operating range of the Mach from 2.6 to 1.9, $\alpha_1(t)$ between -25 to 260, and $\alpha_2(t)$ between 0.52 to 1.27, during the 10 seconds sinusoidal AOA maneuvering, and in the presence of $\pm 50\%$ parameter variation. This is one of the most significant advantage of the EMA controller.

Case 3: NA Tracking Using AOA State Observer

Since in this case, the NA tracking is achieved via a nonlinear algebraic mapping to the AOA tracking, we show in Figures 4.16 and 4.17 only the NA tracking performance results for a step and sine trajectory, respectively. Each figure shows the NA command, the TVB filtered command and the NA output. While no design effort was explicitly directed to the nonminimum phase behavior of the NA tracking, the performances in both cases are remarkable. It was noted during the simulation studies that the NA tracking is very sensitive to variations in the acceleration (curvature) of the tracking command. Thus the performance can be further improved by fine tuning the time-varying bandwidth command logic as given by (3.7), which causes some curvature fluctuation in the filtered command.

Case 4: NA Tracking Using AOA Inner-Loop

As mentioned earlier, this case is currently being studied, so only some preliminary results are shown in Figures 4.18 and 4.19 for step and sinusoidal NA command tracking performance. It is noted that the amplitude of the commands are smaller than that in the previous case, because large command amplitude caused instability. Also in Figure 4.18 a noticeable steady state tracking error is observed, but it indeed converges to zero very slowly, as expected from a Type I system. The sine command tracking result in Figure 4.19 also shows a significant delay. All these problems are due to the intrinsic properties of the integral control, and the limitation of the LTI PI controller in the (nonminimum phase, time-varying) outer loop. Remedies are being investigated at the time of writing.

5. Summary and Conclusions

In this report we have presented the design and simulation study of a missile angle of attack and normal acceleration tracking autopilot using a recently developed extended-mean assignment (EMA) control technique. A radical departure from the conventional design philosophy is that nonlinearity and time-variance of the dynamical systems are not treated as nuances. They are exploited purposely to accomplish design objectives beyond the reach of linear time-invariant control techniques. Salient features of the EMA tracking controller include: (i) good tracking performance for arbitrary trajectories without any scheduling of the constant design parameters throughout the entire operating range of the Mach, (ii) implementation of the inverse pitch dynamics using a static neural network, (iii) time-varying EMA control gains to improve tracking performance, and (vi) a time-varying bandwidth command shaping filter that effectively reduces the actuator rate while maintaining good tracking response for both smooth and abrupt trajectories. Simulation results have shown that the EMA control technique, though still in its embryonic stage, has become a viable design tool for realistic control problems.

Further studies are planned to: (i) implement on-line training of the neural network based inverse plant model to improve tracking accuracy and robustness, (ii) fine tune the time-varying bandwidth command shaping filter to improve the tracking transient performance, (iii) improve the performance of normal acceleration tracking using the angle of attack tracking as inner loop, and (iv) implement a dynamic neural network based "pseudo-inverse" of nonminimum phase plants for arbitrary trajectory tracking. It is noted that this research is only an initial effort to apply the new unified spectral theory and the EMA control technique for LTV systems to practical control problems such as missile autopilot design. The autopilot designed herein is limited to plannar maneuvering only. Design of higher order EMA controllers for multivariable, higher degrees of freedom autopilot is significantly more challenging, and is planned as a long term research goal. Exploring other forms of controllers utilizing the time-varying SD- and PD-eigenvalues is also a long term research goal.
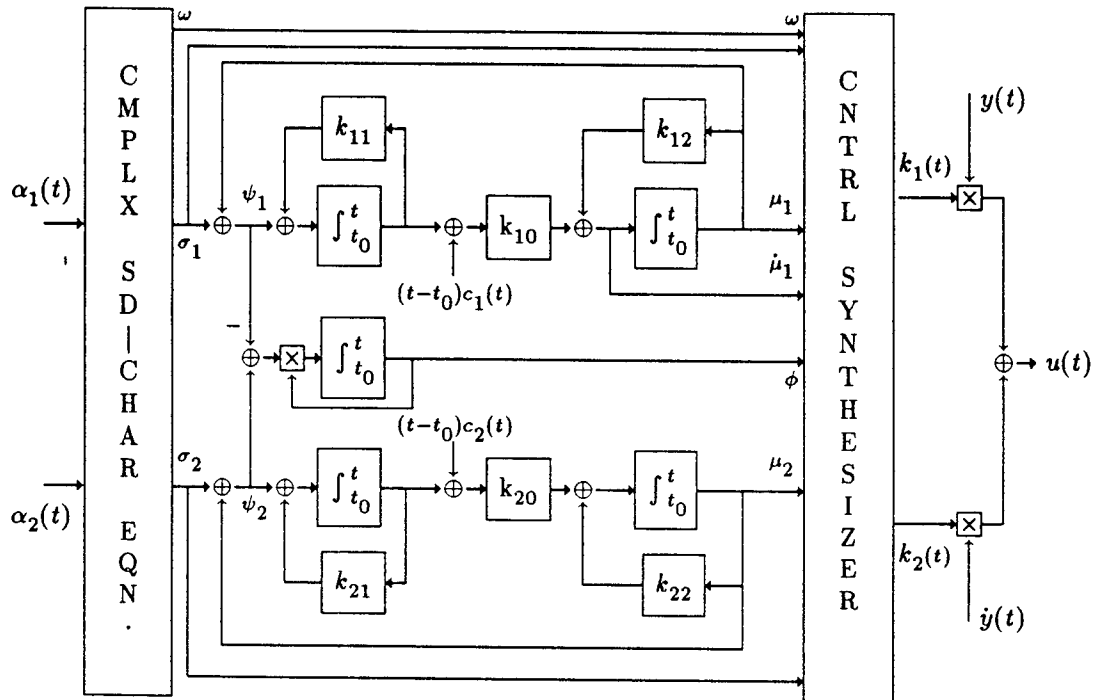
Figure 2.1 The complex-valued EMA controller
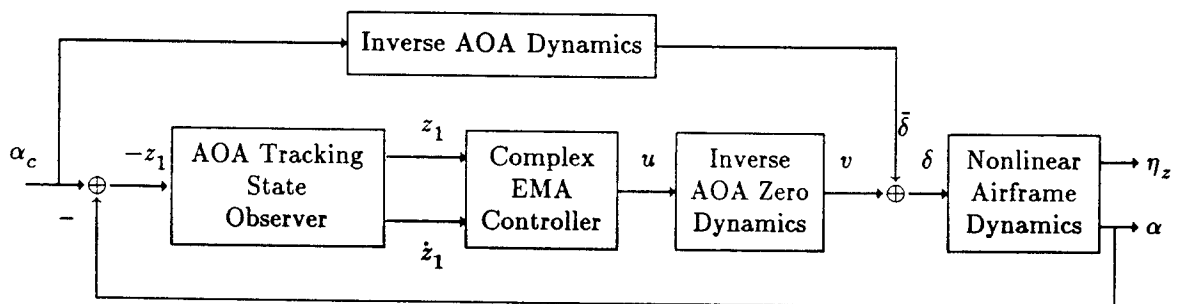
Figure 3.1 AOA Tracking Subsystem
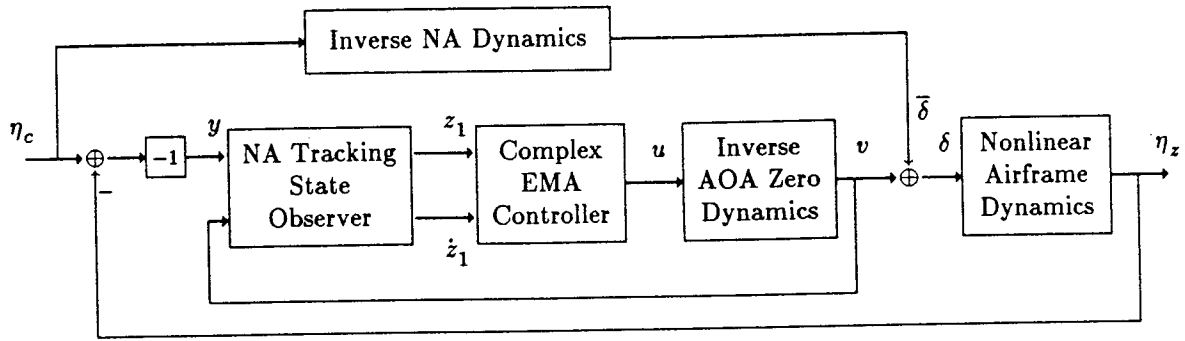
Figure 3.2 NA Tracking System Using AOA State Observer
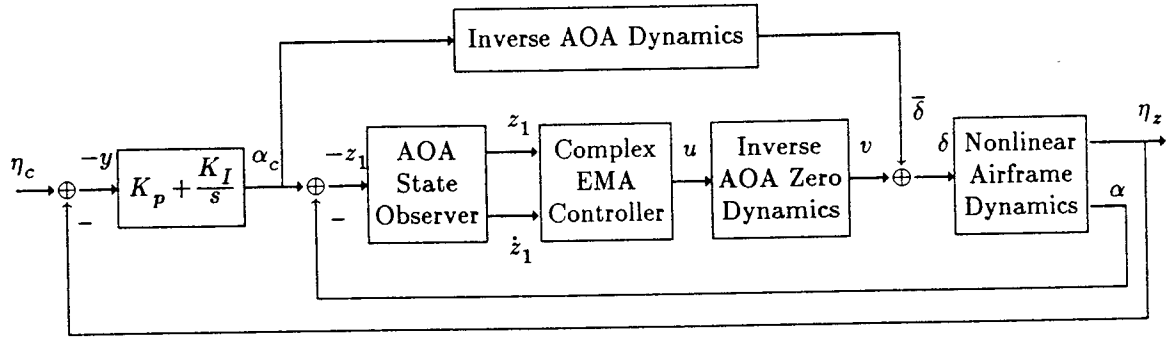
Figure 3.3 NA Tracking System Using AOA Inner-Loop
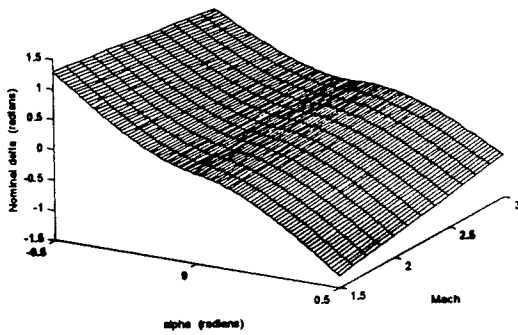
Figure 3.4 The Desired $\overline{\alpha} \mapsto \overline{\delta}$ Mapping
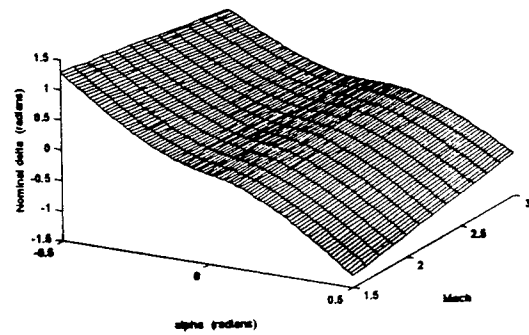
Figure 3.4 The RBF NN $\overline{\alpha} \mapsto \overline{\delta}$ Mapping
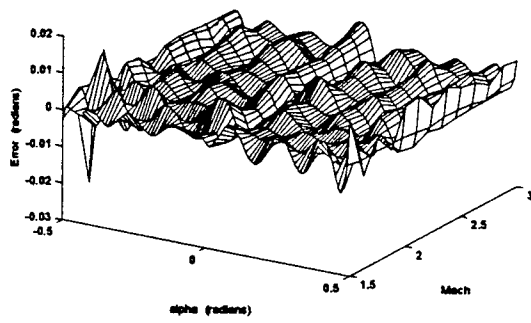
Figure 3.6 Error Surface of the RBF NN $\bar{\alpha} \mapsto \bar{\delta}$ mapping
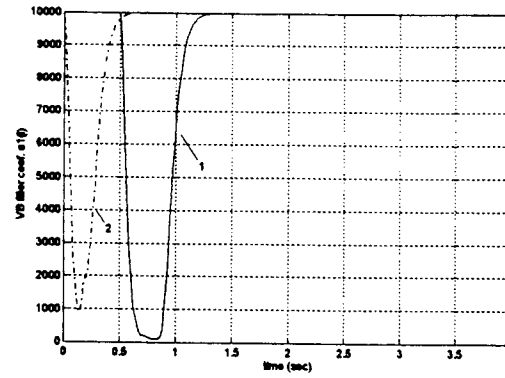
Figure 3.7 Time-varying TVB filter coefficient $a_1(t)$
1—Step command, 2—Sine command
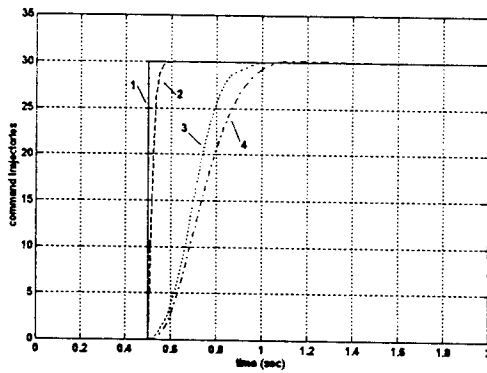
Figure 3.5 TVB Filter Step Response
1—Command, 2—2nd-order LTI filter
3—TVB filter, 4—3rd-order Bessel filter
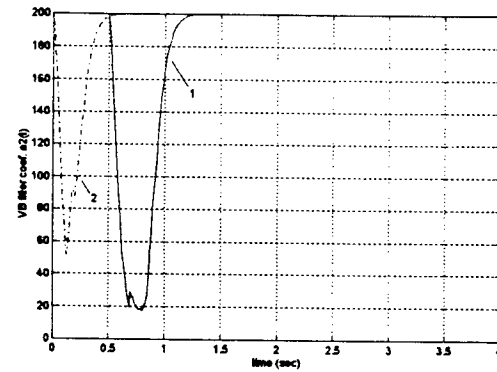
Figure 3.8 Time-varying TVB filter coefficient $a_2(t)$
1—Step command, 2—Sine command
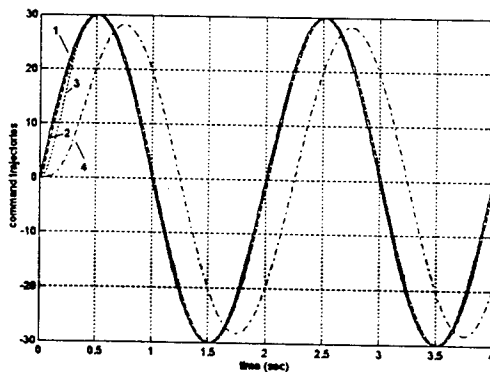
Figure 3.6 TVB Filter Sine Response
1—Command, 2—2nd-order LTI filter
3—TVB filter, 4—3rd-order Bessel filter
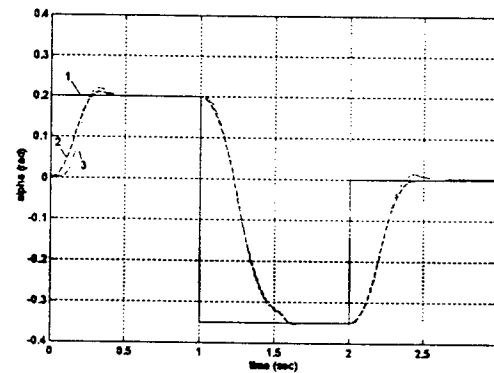
Figure 4.1 Step Trajectory Tracking Performance
1—AOA command, 2—TVB filtered AOA command
3—AOA output
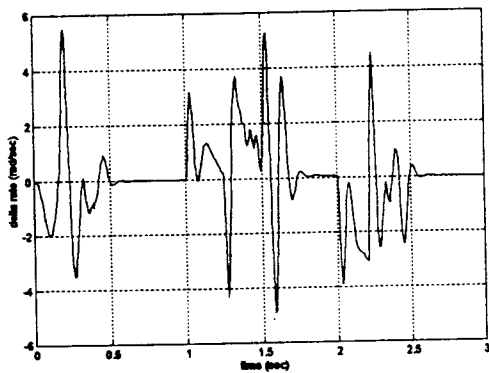
Figure 4.2 Actuator Rate
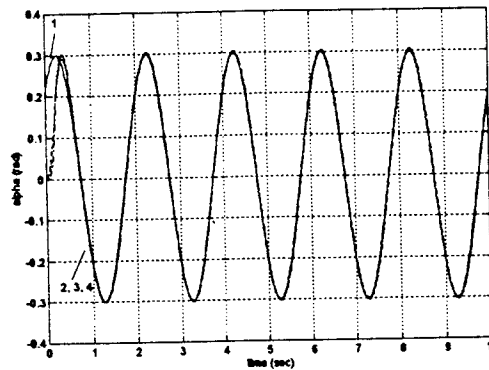
Figure 4.5 Sine Trajectory Tracking Performance
1—AOA command, 2—TVB filtered AOA command
3—Result for Const. EMAC, 4—Result for variable EMAC
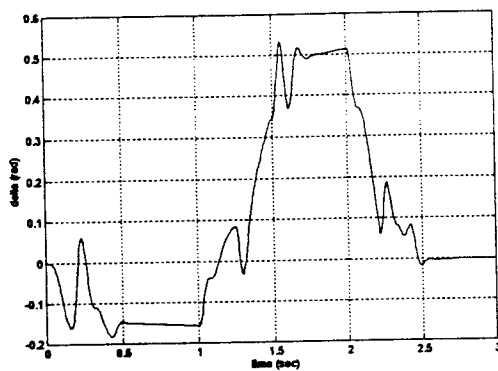
Figure 4.3 Actuator Output
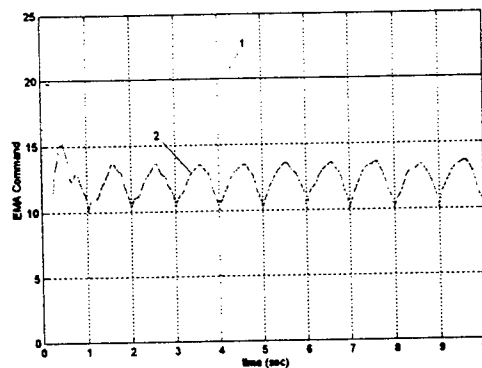
Figure 4.6 Constant vs. Variable EMA Command
1—Constant EMA command, 2—Variable EMA command
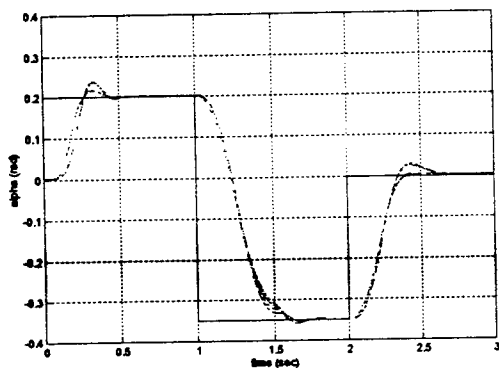
Figure 4.4 Robustness Test — ± 50% variation on C_m, C_n
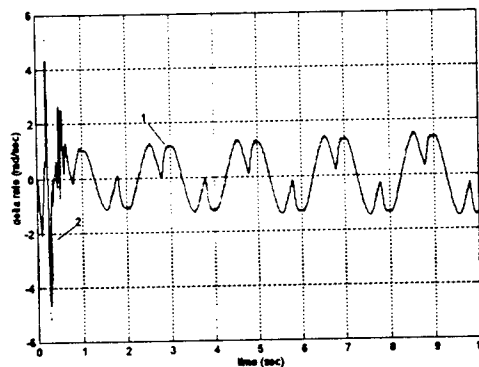
Figure 4.7 Actuator Rate
1—Result for const. EMAC, 2—Result for variable EMAC
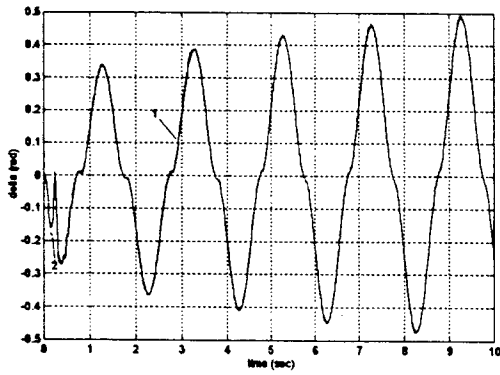
Figure 4.8 Actuator Output
1—Result for const. EMAC, 2—Result for variable EMAC
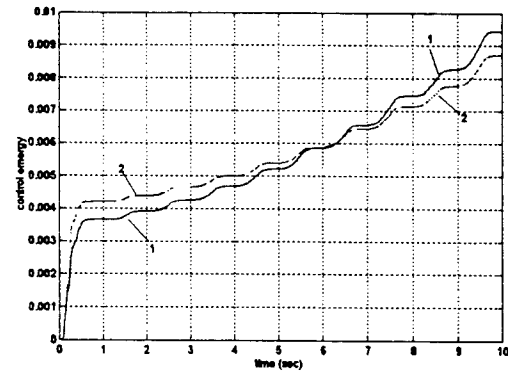
Figure 4.11 Tracking Error Control Energy
1—Result for const. EMAC, 2—Result for variable EMAC
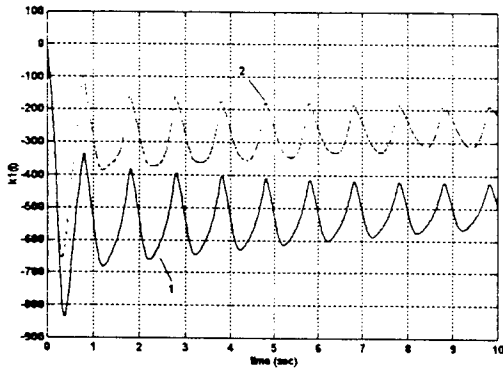
Figure 4.9 Feedback Gain $k_1(t)$
1—Result for const. EMAC, 2—Result for variable EMAC
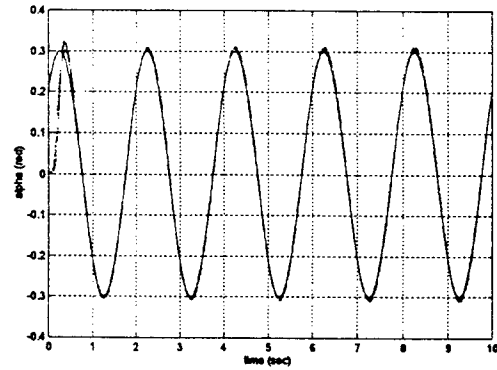
Figure 4.12 Robustness Test
Constant and Variable EMAC, $\pm$ 50% on C_m, C_n
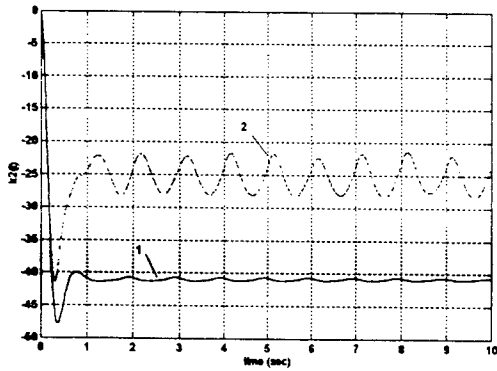
Figure 4.10 Feedback Gain $k_2(t)$
1—Result for const. EMAC, 2—Result for variable EMAC
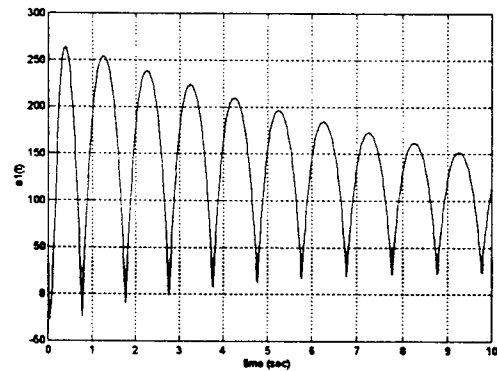
Figure 4.13 Plant Coefficient $\alpha_1(t)$ — Constant EMAC
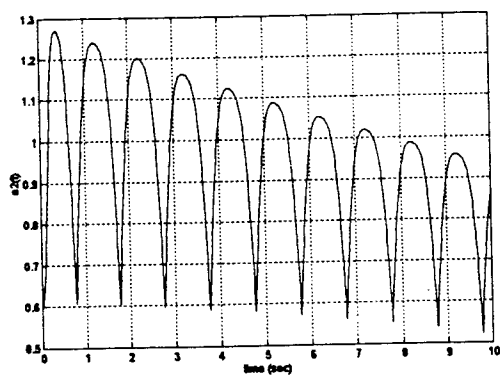
Figure 4.14 Plant Coefficient $\alpha_2(t)$ — Constant EMAC
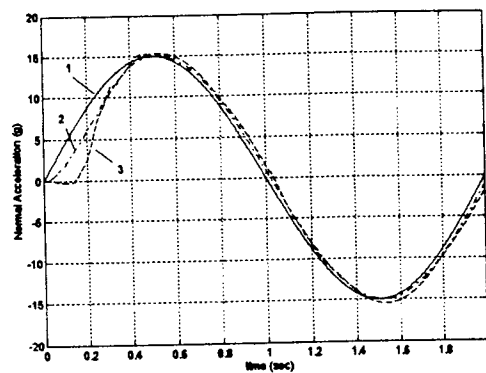
Figure 4.17 NA Sine Trajectory Tracking Performance
— AOA observer
1—NA command, 2—TVB filtered NA com. 3—NA output
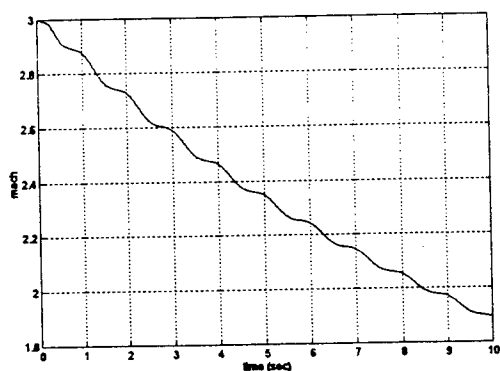
Figure 4.15 Mach Profile — Constant EMAC
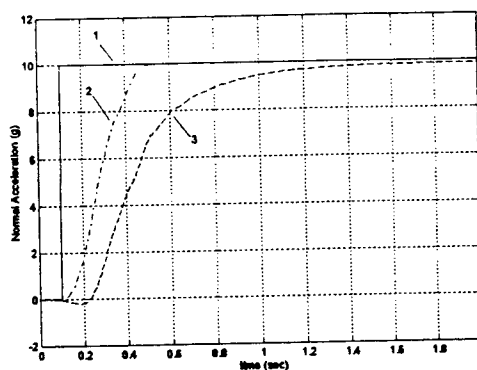
Figure 4.18 NA Step Trajectory Tracking Performance
— AOA inner-loop
1—NA command, 2—TVB filtered NA com. 3—NA output
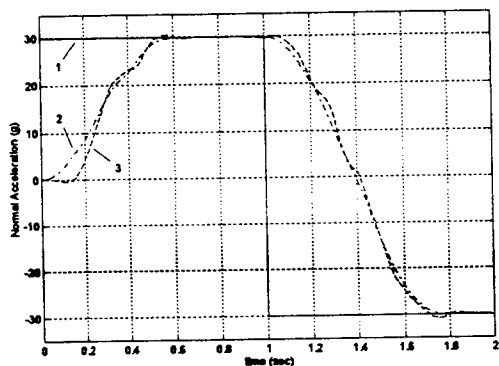
Figure 4.16 NA Step Trajectory Tracking Performance
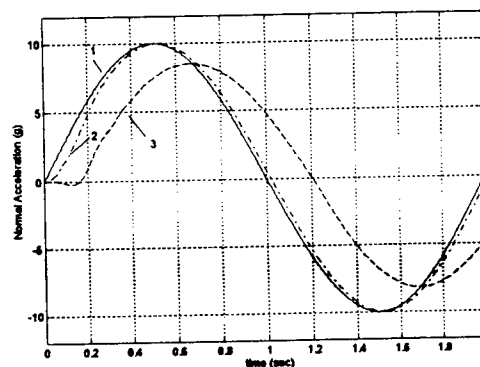— AOA observer
1—NA command, 2—TVB filtered NA com. 3—NA output

Figure 4.19 NA Sine Trajectory Tracking Performance
— AOA inner-loop
1—NA command, 2—TVB filtered NA com., 3—NA output

References

[1] Zhu, J. and Mickle, M. C., "Missile Autopilot Design Using the Extended-Mean Assignment Control, Part I: Stability," *Proc., 27th IEEE SSST*, 247-251, March 1995.

[2] Zhu, J. *"Well-defined Series and Parallel D-Spectra for Linear Time-Varying Systems,"* Proc. Amer. Control Conference, 734-738, Baltimore, MD, June, 1994.

[3] Zhu, J. and Johnson, C. D. "Unified Canonical Forms for Matrices Over a Differential Ring," *Linear Algebra and Its Appl.*, Vol. 147, 201-248, March 1991.

[4] Zhu, J. and Xiao, W., "Intelligent Control of Time-Varying Dynamical Systems Using CMAC Artificial Neural Network," *Mathematical and Computer Modeling*, Special Issue on Neural Networks, Vol. 21, No. 1/2, 89-107, 1995.

[5] White, D. P., Wozniak, J. G. and Lawrence, D. A., "Missile Autopilot Design Using a Gain Scheduling Technique," *Proc., 26th IEEE SSST*, 606-610, March, 1994.

[6] Lawrence, D. A. and Rugh, W. J., "Gain Scheduling Dynamic Linear Controllers for a Nonlinear Plant," *Proc. the 32nd IEEE CDC*, 1024-1029, Dec. 1993.

[7] Nichols, R. A., Reichert, R. T. Rugh W. J., "Gain Scheduling for H-Infinity Controllers: A Flight Control Example," *IEEE Trans. on Control Systems Technology*, Vol. 1, No. 2, 69-79, 1993.

[8] Gopalswamy, S. and Hedrick, J. K., "Control of a High Performance Aircraft with Unacceptable Zerodynamics," Proc. 1992 ACC, 1834-1838, June, 1992.

Acknowledgment: